

# **CRYOGENIC OPERATION OF SILICON-GERMANIUM HETEROJUNCTION BIPOLAR TRANSISTORS AND ITS RELATION TO SCALING AND OPTIMIZATION**

A Thesis  
Presented to  
The Academic Faculty

by

Jiahui Yuan

In Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy in the  
School of Electrical and Computer Engineering

Georgia Institute of Technology  
May, 2010

# CRYOGENIC OPERATION OF SILICON-GERMANIUM HETEROJUNCTION BIPOLAR TRANSISTORS AND ITS RELATION TO SCALING AND OPTIMIZATION

Approved by:

Professor John D. Cressler, Advisor  
School of Electrical and Computer  
Engineering  
*Georgia Institute of Technology*

Professor Shyh-Chiang Shen  
School of Electrical and Computer  
Engineering  
*Georgia Institute of Technology*

Professor Ioannis Papapolymerou  
School of Electrical and Computer  
Engineering  
*Georgia Institute of Technology*

Professor Paul G. Steffes  
School of Electrical and Computer  
Engineering  
*Georgia Institute of Technology*

Dr. Hao Min Zhou  
School of Mathematics  
*Georgia Institute of Technology*

Date Approved: February, 2010

*To my beloved family*

## ACKNOWLEDGEMENTS

I am deeply indebted to Professor John Cressler, with whom I have worked for the past five years. I am extremely fortunate to have had the opportunity to work with such a perfect advisor and mentor.

I would like to express my gratitude to other members of my thesis committee, Professor Shyh-Chiang Shen, Professor Ioannis Papapolymerou, Professor Paul Steffes, and Professor Hao Min Zhou for all their valuable comments and feedback regarding my thesis. In addition, this work could not have been accomplished without the support from Professor Guofu Niu of Auburn University, Professor P. Douglas Yoder of Georgia Tech, and Professor Jae-Sung Rieh of Korea University. The time taken out of their busy schedules for technical discussions is highly appreciated. Support from the industry is essential, and I would like to thank Dr. Alvin Joseph at the IBM System and Technology Group, Dr. Marwan Khater at the IBM T. J. Watson Research Center, and Dr. Holger Rucker at IHP for their technological support as well as insightful suggestions.

I would like to extend my gratitude to Dr. Qingqing Liang, Dr. Enhai Zhao, Dr. Ramkumar Krithivasan, Dr. Joseph Bardin, and Lan Luo for our memorable collaborations and exchange of ideas throughout the past few years. I am also grateful for the support from all other former and current members of Professor Cressler's SiGe group at Georgia Tech, Professor Niu's group at Auburn University, and Professor Weinreb's group at Caltech. They are Professor G. Prakash, Dr. B. Jun, Dr. W.-M. L. Kuo, Dr. L. Najafizadeh, Dr. A. K. Sutton, Dr. C. M. Grens, Dr. T. Chen, Dr. J. Comeau, Dr. C. Zhu, Dr. Y. Lu, Dr. X. Li, Dr. J. Andrews, Dr. M. Bellini, Dr. A. Appaswamy, T. Cheng, T. Thrivikraman, S. Finn, A. Madan, N. Jha, K. Moen, P. Saha, S. Horst, P. Chakraborty, D. Howard, R. Diestelhorst, T. Wilcox, S. Phillips, D. Thomas, S. Seth, J. Poh, T. England, S. Shankar, C. Coen, R. Verma, B. Lieblick,



A. Cardoso, G. Espinel, C. Johnson, S. Huang, Dr. Y. Cui, X. Wei, and H. Mani.

This work was mainly supported by NASA ETDP and SRC. I am grateful for the award selection committee of 2008 *IEEE Electron Device Society Ph.D. Student Fellowship Awards* for the distinguished honor and the financial support.

Most importantly, family means everything to me. I want to thank my parents: Professor Jie Yuan and Peilin Li, and my sister, Dr. Qingqing Yuan, who are always there for me. I sincerely thank my beautiful wife, Yueping Liu, who has been my best friend ever since high school.

# TABLE OF CONTENTS

DEDICATION . . . . .	iii
ACKNOWLEDGEMENTS . . . . .	iv
LIST OF TABLES . . . . .	x
LIST OF FIGURES . . . . .	xi
SUMMARY . . . . .	xviii
<b>I INTRODUCTION . . . . .</b>	<b>1</b>
1.1 Silicon Germanium Heterojunction Bipolar Transistors . . . . .	1
1.2 SiGe HBTs for Cryogenic Applications . . . . .	3
1.2.1 Space Missions at Cryogenic Ambient Temperatures . . . . .	3
1.2.2 Cryogenic Performance of SiGe HBTs . . . . .	5
1.2.3 Heterojunction Barrier Effect . . . . .	7
1.3 SiGe HBTs for Terahertz Systems . . . . .	8
1.3.1 Terahertz Systems . . . . .	8
1.3.2 Terahertz Semiconductor Transistors . . . . .	9
1.3.3 ECL/CML Ring Oscillator Gate Delay . . . . .	11
1.3.4 Vertical and Lateral Scaling in SiGe HBTs . . . . .	12
1.3.5 Non-Equilibrium Base Transport . . . . .	13
1.4 Negative Differential Resistance . . . . .	14
1.5 Higher Breakdown Voltage Design . . . . .	15
1.5.1 Speed / Breakdown Voltage Trade-Off . . . . .	15
1.5.2 History of Superjunctions . . . . .	16
1.6 From Drift-Diffusion to Monte Carlo . . . . .	17
<b>II NOVEL NEGATIVE DIFFERENTIAL RESISTANCE EFFECT AND COLLECTOR-CURRENT KINK EFFECT . . . . .</b>	<b>21</b>
2.1 Novel Experimental Effects . . . . .	21
2.1.1 Negative Differential Resistance Effect . . . . .	21

2.1.2	Collector-Current Kink Effect . . . . .	25
2.2	Review of the Heterojunction Barrier Effect . . . . .	29
2.3	Enhanced Positive Feedback at Deep Cryogenic Temperatures . . .	31
2.4	TCAD Simulation Results . . . . .	34
2.4.1	Collector-Current Kink Effect . . . . .	35
2.4.2	The NDR Picture . . . . .	37
2.5	AC Consequences . . . . .	40
2.6	Collector Doping and Technology Scaling . . . . .	41
2.7	Circuit Implications . . . . .	45
2.8	Summary . . . . .	51
III	BASE PROFILE OPTIMIZATION FOR RF LUNAR APPLICATIONS	52
3.1	Base Carrier Freezeout . . . . .	52
3.2	Base Profile Design . . . . .	55
3.3	Measurement Results . . . . .	56
3.3.1	RF Improvements . . . . .	57
3.3.2	DC Characteristics . . . . .	59
3.4	Physical Origins . . . . .	61
3.5	Summary . . . . .	65
IV	PERFORMANCE LIMITS . . . . .	68
4.1	Scaling and Cooling of SiGe HBTs . . . . .	69
4.1.1	Bandgap Engineering and Vertical Scaling . . . . .	69
4.1.2	Reduction of Parasitics and Charging Times . . . . .	70
4.1.3	Noise . . . . .	71
4.2	Device Optimization for Half-THz Operation . . . . .	73
4.2.1	DC Characteristics . . . . .	74
4.2.2	AC Characteristics . . . . .	78
4.2.3	Cryogenic Performance Limits . . . . .	80
4.3	Gate Delay of Cryogenic Ring Oscillators . . . . .	82

4.3.1	Cryogenic Ring Oscillator Performance . . . . .	83
4.3.2	Self-Heating Across Temperature . . . . .	84
4.4	Temperature Dependence of Intrinsic Base Resistance with Scaling	89
4.5	Summary . . . . .	92
V	SCALING FOR TERAHERTZ SPEEDS . . . . .	94
5.1	The $f_T$ -Doubler Technique . . . . .	94
5.1.1	$f_T$ -Doubler Analysis . . . . .	95
5.1.2	DC and AC Characteristics . . . . .	96
5.1.3	Small-Signal Equivalent Circuit . . . . .	98
5.1.4	Emitter Length Dependence . . . . .	100
5.1.5	Temperature Effects . . . . .	102
5.2	New Scaling Rule and Roadmap . . . . .	102
5.3	New Device Physics For THz SiGe HBTs . . . . .	109
5.3.1	Non-Equilibrium Base Transport . . . . .	109
5.3.2	Base Recombination and NDR . . . . .	110
5.4	TCAD Study of High-Speed SiGe HBTs . . . . .	112
5.4.1	TCAD Calibration . . . . .	113
5.4.2	Scaling Examples . . . . .	113
5.5	Superjunction Collector Design for Improving Breakdown Voltage .	115
5.5.1	Abrupt Superjunction Design . . . . .	116
5.5.2	Practical Profile Design . . . . .	122
5.5.3	Benchmarking . . . . .	124
5.5.4	Comparison with Other Inventions . . . . .	127
5.6	Summary . . . . .	129
VI	MONTE CARLO SIMULATION . . . . .	133
6.1	Issues with TCAD Tools . . . . .	133
6.2	Velocity Overshoot . . . . .	136
6.3	Impact Ionization . . . . .	139

6.3.1	Limitations of Hydrodynamic Models . . . . .	139
6.3.2	Sentaurus Monte Carlo Simulation . . . . .	141
6.3.3	Zener Tunneling . . . . .	143
6.4	Summary . . . . .	144
VII	CONCLUSION AND FUTURE WORK . . . . .	145
	REFERENCES . . . . .	147
	VITA . . . . .	163

## LIST OF TABLES

1	List of key parameters of IBM SiGe HBTs . . . . .	3
2	Comparison of <i>PVCR</i> in SiGe HBTs and in different Si resonant tunneling diodes . . . . .	25
3	Key parameters of the fourth-generation prototype #3 SiGe HBT . .	84
4	Small-signal parameters for a $0.12 \times 2.5 \mu m^2$ third-generation SiGe HBT and an $f_T$ -doubler biased at peak $f_T$ . . . . .	106
5	New SiGe HBT scaling rule and roadmap . . . . .	106
6	Key device parameters of IBM 8HP (130 nm) and SiGe HBTs simulated at Georgia Tech (90 to 45 nm). . . . .	116

## LIST OF FIGURES

1	Architecture of a lunar robotic system using SiGe mixed-signal circuits (Courtesy of Dr. John D. Cressler). . . . .	4
2	Illustration of a lunar surface communication system. . . . .	5
3	The $BV_{CEO}$ - $f_T$ plane of all generations of SiGe HBTs and several examples of record-setting InP-based HBTs, with $f_T \times BV_{CEO}$ contours indicated. . . . .	11
4	Comparison of the measured forced- $V_{BE}$ output characteristics of the SiGe HBT and the control Si BJT at 43 K. . . . .	22
5	Measured forced- $V_{BE}$ output characteristics of a SiGe HBT at (a) 5.4 K, (b) 43 K, (c) 93 K, and (d) 162 K. NDR is observed in (a)–(c). . . . .	23
6	The excess current for NDR (defined as the $I_C$ difference before and after the NDR occurs) in the forced- $V_{BE}$ output characteristics as a function of collector current at 5.4 K, 43 K and 93 K. Inset shows the linear fit for the excess current ratio (defined as the excess current divided by $I_C$ before the NDR) across temperatures. . . . .	24
7	Measured Gummel characteristics of the SiGe HBT at 43 K, 93 K, and 162 K. The $I_C$ kink effect can be observed at 93 K and below, together with the sudden $I_B$ increase representing the onset of heterojunction barrier effect (HBE). . . . .	26
8	Measured Gummel characteristics (with currents on a linear scale) of the SiGe HBT at 43 K and at different $V_{CB}$ 's. Also shown is the transistor transconductance ( $g_m$ ) at $V_{CB} = 0$ V. Points A and B correspond to points A and B in Fig. 4, indicating the intrinsic relationship between NDR and the kink effect. . . . .	27
9	Measured Gummel characteristics of the control Si BJT at 43 K and at $V_{CB} = 0$ V and $V_{CB} = -0.5$ V. Also shown is the transistor transconductance ( $g_m$ ) at $V_{CB} = 0$ V. . . . .	28
10	Simulated conduction band edge ( $E_C$ ) and electron density ( $n$ ) as functions of depth for the SiGe HBT at 77 K, at 300 K and the control Si BJT at 77 K. The $V_{BE}$ 's are chosen to ensure comparable collector current for all the three situations. Positions for the metallurgical junctions and the cut line are marked. . . . .	32
11	The feedback loop of the key physical parameters happening at heterojunction barrier effect and at (a) higher temperatures forming a negative feedback, (b) deep cryogenic temperatures forming a positive feedback. . . . .	33

12	DESSIS simulations of the SiGe HBT at 77 K, including (a) forced- $V_{BE}$ output characteristics showing NDR, and (b) Gummel characteristics and the resultant $g_m$ where the $I_C$ kink effect occurs. . . . .	34
13	DESSIS simulations of the forced- $V_{BE}$ output characteristics of the SiGe HBT at 300 K, and the control Si BJT at 77 K. . . . .	35
14	Simulated heterojunction barrier height as functions of $V_{CE}$ for the SiGe HBT at (a) 300 K ( $V_{CB} = 0.0$ V) and (b) 77 K ( $V_{CB} = 0.0$ and 1.0 V) . . . . .	36
15	Simulated electron density as a function of $V_{BE}$ for the SiGe HBT at (a) 300 K ( $V_{CB} = 0.0$ V) and (b) 77 K ( $V_{CB} = 0.0$ and 1.0 V). The snapshot is taken at the position of the cut line indicated in Fig. 10. .	37
16	Simulated heterojunction barrier height as a function of $V_{CE}$ for the SiGe HBT at 77 K and 300 K. The snapshot is taken at the position of the cut line indicated in Fig. 10. . . . .	38
17	Simulated electron density and electric field as functions of $V_{CE}$ for the SiGe HBT at 77 K and 300 K. The snapshot is taken at the position of the cut line indicated in Fig. 10. . . . .	39
18	Simulated drift component of the electron current density and the total electron current density as functions of $V_{CE}$ for the SiGe HBT at 77 K and 300 K. The snapshot is taken at the position of the cut line indicated in Fig. 10. . . . .	40
19	Measured cut-off frequency ( $f_T$ ) <i>vs.</i> collector current ( $I_C$ ) for the SiGe HBT at 77 K and 300 K. The 77 K plot shows an $f_T$ “dip” in accordance with the $I_C$ kink effect in its DC characteristics. . . . .	41
20	Measured Gummel characteristics and transistor transconductance ( $g_m$ ) at $V_{CB} = 0$ V for the HB and HP first-generation SiGe HBTs at 43 K. . . . .	43
21	Measured Gummel characteristics and transistor transconductance ( $g_m$ ) at $V_{CB} = 0$ V for (a) the third-generation SiGe HBT (200 GHz at 300 K) at 43 K, and (b) the fourth-generation SiGe HBT (350 GHz at 300 K) at 4.3 K. . . . .	44
22	Measured peak unit transconductance ( $g'_m$ ) as a function of $V_{CB}$ for the first- (both HP and HB), third-, and fourth-generation SiGe HBTs. . . . .	45
23	Comparison of the different NDR effects in the first-generation and the state-of-the-art fourth generation SiGe HBTs. NDR is observed in the forced- $V_{BE}$ output characteristics of the first-generation SiGe HBTs, but in the forced- $I_B$ output characteristics of the fourth-generation devices together with a hysteresis behavior. . . . .	46



24	The measured forced- $V_{BE}$ output characteristics of SiGe HBT at 43 K and at $V_{BS} = 1.16$ V by changing the value of the base series resistance.	49
25	The measured excess current ratio at 43 K as a function of the base series resistance. . . . .	50
26	Temperature dependence of the normalized base sheet resistance of a first-generation SiGe HBT in comparison with the normalized resistance of other p-type Si/SiGe layers (data provided by K. Moen). . .	54
27	The doping (in logarithm scale) and Ge (in linear scale) profiles for the control and optimized SiGe HBTs. . . . .	56
28	Forward Gummel characteristics of a SiGe HBT ( <i>new Ge(x) + higher <math>N_B</math></i> ) at different temperatures. . . . .	57
29	Measured $f_T$ and $f_{max}$ vs. $I_C$ for the three profiles at 300 K. . . . .	58
30	Measured $f_T$ and $f_{max}$ vs. $I_C$ for the three profiles at 43 K. . . . .	59
31	Peak $f_T$ and $f_{max}$ vs. temperature for the three profiles. . . . .	60
32	Measured $\beta$ vs. $I_C$ for the three profiles at 300 K. . . . .	61
33	Measured $\beta$ vs. $I_C$ for the three profiles at 43 K. . . . .	62
34	Peak $\beta$ vs. temperature for the three profiles. . . . .	63
35	Measured $g_m$ vs. $I_C$ for the three profiles at 43 K and 300 K. . . . .	64
36	Measured $g_m$ at $I_C = 1.8$ mA vs. temperature for the three profiles. .	65
37	Extracted $R_B$ vs. $I_C$ for the three profiles at 300 K and 43 K. . . . .	66
38	Measured $BV_{CEO}$ and $BV_{CBO}$ vs. temperature for the three profiles. .	67
39	$f_T$ and $f_{max}$ as functions of collector current for a second-generation SiGe HBT at 300 K and 93 K. . . . .	72
40	Effective noise temperature (K) and gain (dB) of the X-band LNA at 300 K and 15 K (measured by T. Thirvikraman). . . . .	73
41	Forward Gummel characteristics for a fourth-generation prototype #3 SiGe HBT at 300, 112, and 4.5 K. . . . .	76
42	Current gain at $V_{CB} = 0.5$ V for a $0.12 \times 2.5 \mu m^2$ SiGe HBT (fourth-generation prototype #3) as a function of collector current density, at 300, 223, 162, 112, 77, and 4.5 K. . . . .	77
43	Forced- $I_B$ and forced- $V_{BE}$ output characteristics of a $0.12 \times 2.5 \mu m^2$ SiGe HBT (fourth-generation prototype #3) at 4.5 K. . . . .	78

44	$f_T$ and $f_{max}$ as functions of collector current for a $0.12 \times 2.5 \mu m^2$ SiGe HBT (fourth-generation prototype #3) at 300, 112, and 4.5 K. . . . .	79
45	Small-signal power gain ( $U$ ) and current gain ( $h_{21}$ ) for a $0.12 \times 2.5 \mu m^2$ SiGe HBT (fourth-generation prototype #3) at 300 and 4.5 K. . . . .	80
46	$f_T$ and $f_{max}$ extracted at various frequencies between 15 GHz and 35 GHz based on -20 dB/dec roll-off of $h_{21}$ and $U$ . . . . .	81
47	S-parameters ( $S_{11}$ , $S_{12}$ , $S_{21}/10$ , and $S_{22} - 1$ ) at peak $f_{max}$ for a $0.12 \times 2.5 \mu m^2$ SiGe HBT (fourth-generation prototype #3) at 300, 112, and 4.5 K. . . . .	82
48	Total delay versus reciprocal collector current density at constant $V_{CB}$ for a $0.12 \times 2.5 \mu m^2$ SiGe HBT (fourth-generation prototype #3), at 300, 112, and 4.5 K. . . . .	83
49	Peak $f_T$ and $f_{max}$ as functions of the chuck temperature for all the three prototypes of the fourth-generation SiGe HBTs. . . . .	85
50	Relative location of the present result on the $f_T$ - $f_{max}$ plane in comparison with other reported values of SiGe HBTs. . . . .	86
51	Schematic of (a) the CML ring oscillator, (b) cross-section of the SiGe HBT structure, and (c) current mirror and one CML gate of the ring oscillator. . . . .	87
52	Measured minimum CML ring oscillator gate delay as a function of chuck temperature. . . . .	88
53	Measured CML ring oscillator gate delay <i>vs.</i> current per gate at 25, 77, 170, 294 K. . . . .	89
54	Self-heating-induced temperature increase: $\Delta T_{total} = \Delta T_1 + \Delta T_2$ , where $\Delta T_1$ is the increase of the average temperature in the circuit area ( $T_{RO}$ ) relative to the chuck temperature ( $T_{chuck}$ ), and $\Delta T_2$ is the increase of the HBT temperature ( $T_{HBT}$ ) relative to $T_{RO}$ . . . . .	90
55	Measured temperature dependence of the normalized base sheet resistance of a third-generation SiGe HBT (peak doping about $3 \times 10^{19} cm^{-3}$ ) in comparison with the normalized sheet resistance of other p-type Si/SiGe layers that were shown in Fig. 24. . . . .	91
56	(a) Schematic of the $f_T$ -doubler; (b) small-signal equivalent circuit of a SiGe HBT. . . . .	95
57	Measured forward Gummel characteristics of a third-generation SiGe HBT and a SiGe HBT $f_T$ -doubler, using an emitter geometry of $0.12 \times 2.5 \mu m^2$ . . . . .	97

58	DC current gains ( $\beta$ ) and transconductances ( $g_m$ ) <i>vs.</i> $I_C$ extracted from DC characteristics of the SiGe HBT and the $f_T$ -doubler. . . . .	98
59	Forced- $I_B$ output characteristics of the SiGe HBT and the $f_T$ -doubler.	99
60	$f_T$ and $f_{max}$ as functions of $I_C$ of the SiGe HBT and the $f_T$ -doubler extrapolated at 40 GHz. . . . .	100
61	Maximum available gain ( $MAG$ ) and $h_{21}$ of the $f_T$ -doubler and the SiGe HBT biased at a collector current needed to attain peak $f_T$ . . .	101
62	Extracted $R_b$ and $G_{m0}$ as functions of $I_C$ for the SiGe HBT and the $f_T$ -doubler. . . . .	102
63	Extracted $R_\pi$ and $C_\pi$ as functions of $I_C$ for the SiGe HBT and the $f_T$ -doubler. . . . .	103
64	Extracted $R_O$ and $C_{bc}$ as functions of $I_C$ for the SiGe HBT and the $f_T$ -doubler. . . . .	104
65	Comparison of the measured and modeled S-parameters ( $S_{11}$ , $S_{12}$ , $S_{21}/10$ , and $S_{22}$ ) for the $f_T$ -doubler. . . . .	105
66	Measured peak $f_T$ as a function of emitter length ( $L_E$ ) for SiGe HBTs and $f_T$ -doublers. . . . .	107
67	Measured peak $f_T$ as a function of temperature for the SiGe HBT and the $f_T$ -doubler. . . . .	108
68	Collector current of a fourth-generation prototype #3 SiGe HBT at different temperatures. Inset shows the effective carrier temperature extracted from the slope of $I_C$ as a function of chuck temperature. . .	110
69	Peak transconductance and peak current gain at $V_{CB} = 0.5$ V for a fourth-generation prototype #3 SiGe HBT as a function of temperature. . . . .	111
70	Comparison of measured and simulated $f_T$ and $f_{max}$ <i>vs.</i> $J_C$ of a third-generation SiGe HBT. . . . .	114
71	Comparison of measured and simulated $M - 1$ and $I_B$ <i>vs.</i> $V_{CB}$ of a third-generation SiGe HBT. . . . .	115
72	Cutoff frequency ( $f_T$ ) as a function of $I_C$ for a fourth-generation SiGe HBT and four experimental SiGe HBTs simulated at Georgia Tech. .	117
73	Maximum oscillation frequency ( $f_{max}$ ) as a function of $I_C$ for a fourth-generation SiGe HBT and four experimental SiGe HBTs simulated at Georgia Tech. . . . .	118

74	Simulated electron temperature and electric field in a SiGe HBT with standard collector (the peak $f_T = 105$ GHz and $BV_{CEO} = 2.67$ V ). . . . .	119
75	The doping profile used in the TCAD simulation, including the standard collector doping and the PN superjunction. . . . .	120
76	Comparison of electrostatic potential and electric field of the standard collector with that of a superjunction collector. . . . .	121
77	Comparison of electron temperature and ionization coefficient of the standard collector with that of a superjunction collector. . . . .	122
78	Simulated multiplication factor for the IBM 8WL, standard collector, and superjunction collector SiGe HBTs. . . . .	123
79	Comparison of $f_T$ and $f_{max}$ of the standard collector SiGe HBT with that of a superjunction one. . . . .	124
80	The doping profile of Gaussian superjunction representing the impact of thermal cycles. . . . .	125
81	Simulated multiplication factor for a standard, abrupt superjunction, and Gaussian superjunction collector SiGe HBTs. . . . .	126
82	The doping profile of the simplified superjunction collector SiGe HBT. . . . .	127
83	Output characteristics of the simplified superjunction collector SiGe HBT. . . . .	128
84	$M - 1$ for a standard, abrupt superjunction, and simplified superjunction collector SiGe HBTs. . . . .	129
85	$f_T$ and $f_{max}$ of a standard, superjunction, and simplified superjunction collector SiGe HBTs. . . . .	130
86	$f_T$ <i>vs.</i> $V_{CB}$ of a standard, superjunction, and simplified superjunction collector SiGe HBTs. . . . .	131
87	Benchmarking: peak $f_T$ <i>vs.</i> $BV_{CEO}$ for SiGe HBTs. . . . .	132
88	Simulated electron density <i>vs.</i> depth in a Si BJT with doping profiles equivalent to a 200-GHz SiGe HBT. The dashed line denotes hydrodynamic simulation, and the solid line shows Monte Carlo simulation. . . . .	135
89	Simulated electron temperature (or electron energy) and electric field <i>vs.</i> depth in a Si BJT with doping profiles equivalent to a 200-GHz SiGe HBT. The dashed lines denote hydrodynamic simulation, and the solid lines show Monte Carlo simulation. . . . .	137

90	Simulated electron velocity <i>vs.</i> depth in a Si BJT with doping profiles equivalent to a 200-GHz SiGe HBT. The dashed line denotes hydrodynamic simulation, and the solid line shows Monte Carlo simulation. . . . .	138
91	Simulated electron impact ionization rate <i>vs.</i> depth in a Si BJT with doping profiles equivalent to a 200-GHz SiGe HBT. The dashed line denotes hydrodynamic simulation, and the solid line shows Monte Carlo simulation. . . . .	143

## SUMMARY

The objective of the proposed work is to study the behavior of SiGe HBTs at cryogenic temperatures and its relation to device scaling and optimization. Not only is cryogenic operation of these devices required by space missions, but characterizing their cryogenic behavior also helps to investigate the performance limits of SiGe HBTs and provides essential information for further device scaling. Technology computer aided design (TCAD) and sophisticated on-wafer DC and RF measurements are essential in this research.

Drift-diffusion (DD) theory is used to investigate a novel negative differential resistance (NDR) effect and a collector current kink effect in first-generation SiGe HBTs at deep cryogenic temperatures. A theory of positive feedback due to the enhanced heterojunction barrier effect at deep cryogenic temperatures is proposed to explain such effects. Intricate design of the germanium and base doping profiles can greatly suppress both carrier freezeout and the heterojunction barrier effect, leading to a significant improvement in the DC and RF performance for NASA lunar missions.

Furthermore, cooling is used as a tuning knob to better understand the performance limits of SiGe HBTs. The consequences of cooling SiGe HBTs are in many ways similar to those of combined vertical and lateral device scaling. A case study of low-temperature DC and RF performance of prototype fourth-generation SiGe HBTs is presented. This study summarizes the performance of all three prototypes of these fourth-generation SiGe HBTs within the temperature range of 4.5 to 300 K. Temperature dependence of a fourth-generation SiGe CML gate delay is also examined, leading to record performance of Si-based IC. This work helps to analyze the key optimization issues associated with device scaling to terahertz speeds at room

temperature. As an alternative method, an  $f_T$ -doubler technique is presented as an attempt to reach half-terahertz speeds. In addition, a roadmap for terahertz device scaling is given, and the potential relevant physics associated with future device scaling are examined. Subsequently, a novel superjunction collector design is proposed for higher breakdown voltages. Hydrodynamic models are used for the TCAD studies that complete this part of the work. Finally, Monte Carlo simulations are explored in the analysis of aggressively-scaled SiGe HBTs.

A significant amount of this work has been published or submitted for publication at various refereed conferences and journals, including *IEEE International Electron Device Meeting* [18], *IEEE Transactions on Electron Devices* [19]-[22], *IEEE Bipolar/BiCMOS Circuits and Technology Meeting* [23]-[27], *IEEE Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems* [28], and *IEEE Microwave Theory and Wireless Component Letters* [29]. Master level research prior to this work was completed in 2007 [30]. More specifically, details of this dissertation can be found in the following refereed publications:

1. Discovery, explanation, and simulation of a novel cryogenic negative differential resistance effect and a collector-current kink effect in first-generation SiGe HBTs (Chapter II, also published as [18]).
2. Modeling of the above novel effects and study of their circuit implications (Chapter II, also published as [19]).
3. Base profile optimization for improved RF performance at 43 K (Chapter III, also published as [25]).
4. Record  $f_{max}$  achieved for all silicon-based technologies (Chapter IV, also published as [23]).
5. The fastest CML/ECL gate delay achieved for all silicon-based technologies and the first report of the temperature dependence of intrinsic base sheet resistance in third-generation SiGe HBTs (Chapter IV, also published as [22]).

5. A scaling path for SiGe HBTs explored through its relation to cryogenic operation of prototype fourth-generation SiGe HBTs (Chapter V, also published as [21]).
6. First demonstration of the  $f_T$ -doubler technique in third-generation SiGe HBTs to achieve half-THz performance (Chapter V, also published as [28]).
7. Invention of a superjunction collector design in high-speed bipolar transistors to improve the speed/breakdown tradeoff (Chapter V, also published as [27]).



# CHAPTER I

## INTRODUCTION

Combining the high performance of III-V transistors with the low cost and high yield of conventional silicon (Si) devices, silicon-germanium (SiGe) heterojunction bipolar transistors (HBTs) have achieved great success in modern radio frequency (RF) and millimeter-wave integrated circuit (IC) applications [1]. In addition to its growing importance in these applications, the SiGe HBT is also being pursued for a host of so-called “extreme environment” applications [2]. Such applications involve, for instance, operation under radiation exposure (e.g., for space exploration or Earth orbit) [3], [4], as well as at reduced (cryogenic) temperatures (e.g., to 77 K or even 4 K) [5]-[9]. Note that the surface of the Moon can reach as low as 43 K in the shadowed polar craters. Meanwhile, the performance levels of the fully silicon-manufacturing compatible SiGe HBT are steadily marching upward with generational scaling [10]-[17], making the device a new contender for even higher frequency ranges. The possibility of using highly integrated SiGe IC platforms for emerging communications and terahertz (THz) systems is highly appealing from a cost perspective.

### ***1.1 Silicon Germanium Heterojunction Bipolar Transistors***

The heterojunction (or heterostructure) bipolar transistor was first proposed by W. Shockley in 1948 as a bipolar transistor with a wide-gap emitter [31]. In his patent disclosure, Shockley described his invention as a device “in which one of the separated zones is of semiconductive material having a wider energy gap than that of the material in the other zones.” Despite the great potential of these devices [32], [33], manufacturable HBTs did not exist until the 1970’s when a few epitaxial technologies (e.g. liquid-phase epitaxy (LPE), molecular beam epitaxy (MBE), and metal-organic

chemical vapor deposition (MOCVD)) made the highly advanced HBTs routinely available [34]]. Since then, heterojunction transistors made from III-V materials have achieved a great success [35]-[37].

Such bandgap-engineering was applied to silicon-based transistors in the 1980's, and SiGe HBTs gradually became a strong candidate for RF to millimeter-wave communication ICs. Combining conventional shallow and deep trench isolation with CMOS back-end-of-line (BEOL) processing, one can integrate SiGe HBTs and Si CMOS devices on the same wafer. This so-called SiGe BiCMOS technology is 100% silicon manufacturing compatible [38].

Strictly speaking, a SiGe HBT is a double-heterojunction bipolar transistor (DHBT) as defined in [34]. Both the emitter and collector silicon material have a wider bandgap than that of the SiGe in the epitaxial base. The heterojunction in the emitter-base (EB) junction prevents the back-injection of the holes from base to emitter, thus increasing the current gain ( $\beta$ ); the electric field resulting from the graded bandgap decreases the base transit time ( $\tau_B$ ); the heterojunction at the collector-base (CB) junction also increases the Early voltage ( $V_A$ ) [39]. The performance of SiGe HBTs has seen steady improvement in the past decade [10]-[17]. Now the fourth-generation SiGe HBT has a cutoff frequency ( $f_T$ ) of 300 GHz, a maximum oscillation frequency ( $f_{max}$ ) of 350 GHz, and an open-base emitter-collector breakdown voltage ( $BV_{CEO}$ ) of 1.7 V [15]. Other competitors have recently reported transistors with  $f_{max}$  higher than 400 GHz [41], [17].

Table 1 summarizes the performance of the four generations of IBM SiGe HBTs as measured by the author. Here the technologies 5HP/6HP, 7HP, 8HP, and 9T are typical representatives of the first-, second-, third-, and fourth-generation SiGe HBTs, and technological details can be found in [40] and [15].

**Table 1:** List of key parameters of IBM SiGe HBTs

Parameters	5HP	6HP	7HP	8HP	9T
Lithographic node ( $\mu m$ )	0.5	0.25	0.18	0.13	0.13
$f_T$ (GHz)	47	47	120	210	300
$f_{max}$ (GHz)	65	65	100	285	350
$\beta$	100	100	350	300	650
$BV_{CEO}$ (V)	3.4	3.4	1.8	1.7	1.7
$BV_{CBO}$ (V)	10.5	10.5	6.5	5.5	5.6
$J_C$ @ peak $f_T$ (mA/ $\mu m^2$ )	1.5	1.5	8	12	19

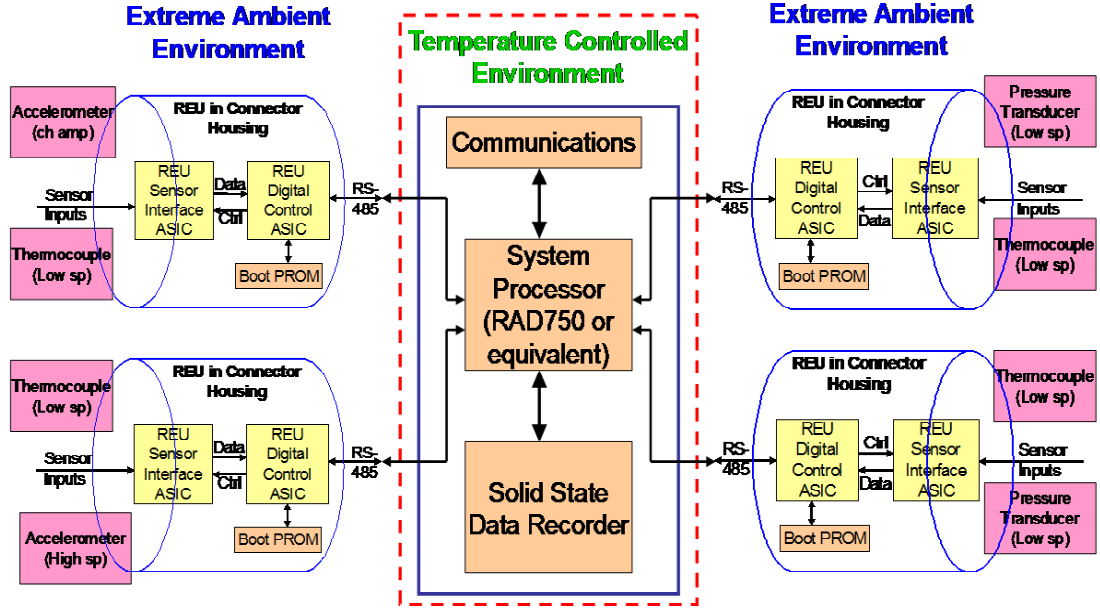
## 1.2 *SiGe HBTs for Cryogenic Applications*

SiGe HBTs are ideal candidates for the emerging cryogenic applications, including the NASA space missions.

### 1.2.1 Space Missions at Cryogenic Ambient Temperatures

Lunar missions require numerous digital, analog, and mixed-signal circuit blocks for data conversion and processing, RF communication circuits, power conditioning, actuation, control, and sensor interfaces. These are mostly mixed-signal circuits that require excellent reliability and immunity to the high radiation levels in extraterrestrial environments.

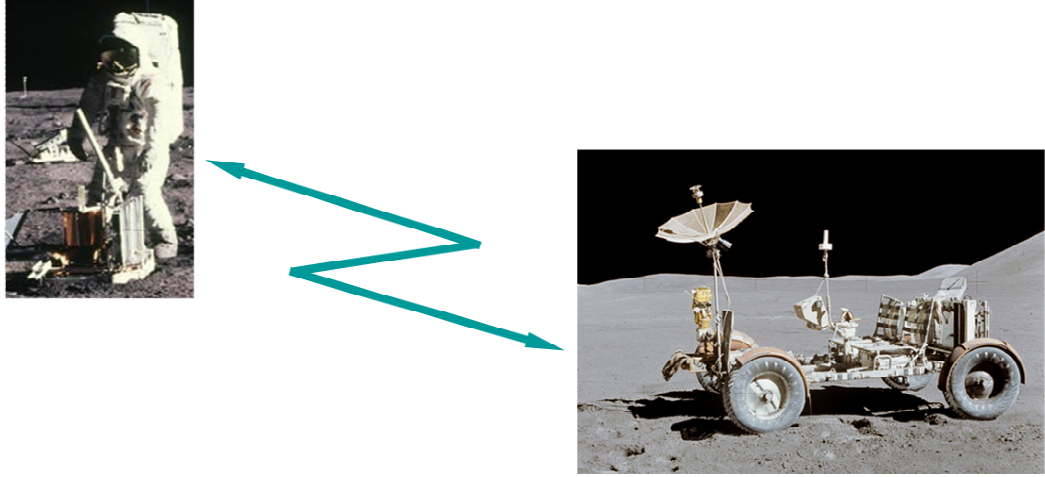
Fig. 1 shows an ideal lunar robotic system on a rover that has electronics, sensors, and actuators for control. The distributed sensor and actuator networks monitor the health and performance of the rover. One of the major challenges for this system is the extreme temperature conditions on the lunar surface, where the temperature ranges from -180C (93 K) at lunar night to +120C (393 K) during the day. The shadowed polar craters can even reach -230C, precluding the use of conventional terrestrial electronics for sensing, actuation, and control under ambient conditions. The conventional method of preventing the electronic system from being damaged



**Figure 1:** Architecture of a lunar robotic system using SiGe mixed-signal circuits (Courtesy of Dr. John D. Cressler).

by the cold environment is to use a “warm box” that is heavy, power hungry, and results in reliability issues [2]. The distributive nature of the sensing networks causes serious wiring and placement issues within the “warm box”. Therefore, semiconductor ICs that can operate at the extreme ambient environment are highly desirable for eliminating the need for a “warm box”.

Another promising application is the lunar surface communication system (Fig. 2), which enables astronauts to communicate with each other or the rovers through RF wireless phones. Again, these devices have to operate at the extreme ambient environment to ensure the mobility of the dispersed communication system. In addition to the reliability requirement, semiconductor devices for such applications need to provide consistent RF performance across the entire temperature range. However, the majority of semiconductor transistors currently available on the market cannot meet this requirement.



**Figure 2:** Illustration of a lunar surface communication system.

### 1.2.2 Cryogenic Performance of SiGe HBTs

Unlike conventional Si BJTs, whose cryogenic performance is limited by the heavy doping-induced emitter bandgap narrowing and the increased base transit time ( $\tau_B$ ) due to the higher electron diffusivity [39], SiGe technology is a natural fit for such niche applications because it possesses inherent ionizing radiation tolerance as fabricated, and the impact of bandgap engineering on the speed of SiGe HBTs is in general favorably affected by cooling.

SiGe voltage references are now flying in the NASA MISSE-6 mission (on the international space station), and more SiGe electronic sub-systems using first-generation SiGe HBTs for sensing, actuation, and robotics on the Moon are under development as a part of the NASA exploratory technology development program. Meanwhile, another logical application of SiGe would be in cryogenically-enabled wireless S, Ka, and K-band communication systems. Improving the RF performance of first-generation SiGe HBTs below 100 K becomes necessary in this context.

Pioneering work on cooled SiGe HBTs has demonstrated improved performance

at cryogenic temperatures [39]. Nevertheless, such cooling-induced performance enhancements were not dramatic, partly due to the relatively small bandgap grading used in first-generation SiGe HBTs, together with the finite base freezeout associated with the modest base doping [5]. However, cooling has been reported to improve the speed of more advanced SiGe HBTs, partly because it magnifies the favorable effects of bandgap engineering. For instance, the peak  $f_T$  and  $f_{max}$  of the SiGe HBT in reference [41] increase from 253 and 281 GHz at 300 K to 404 and 443 GHz at 50 K. The highest  $f_T$  reported to-date has exceeded 600 GHz at deep cryogenic temperatures [42]. Our group has recently reported the first half-terahertz SiGe HBT (a prototype fourth-generation SiGe HBT) reaching  $f_T = 510$  GHz,  $f_{max} = 276$  GHz, and  $BV_{CEO} = 1.36$  V at 4.5 K ( $f_T = 352$  GHz,  $f_{max} = 241$  GHz, and  $BV_{CEO} = 1.47$  V at 300 K) [43]. Based on these initial results, further optimization was made to achieve a more balanced  $f_T/f_{max}$  of 463 / 618 GHz at 4.5 K, as reported in [23]. This optimization leads to a scaling path that will be discussed in more detail in this work. Thus, cooling can be used as an effective way to better understand the ultimate scaling limits of SiGe HBTs.

Clearly, however, a detailed understanding of the subtleties of device operation under these extreme temperatures is required if circuit applications are to follow. Unusual phenomena can indeed be found in SiGe HBTs, especially at very low temperatures. Trap-assisted tunneling, for example, induces a minority carrier transport mechanism that can produce a non-ideal collector current component at temperatures below 77 K [44]. In addition, an I-V hysteresis and negative differential resistance (NDR) were recently observed in the forced- $I_B$  output characteristics of aggressively-scaled fourth-generation SiGe HBTs, and they were shown to be the result of enhanced tunneling and recombination at high injection and at cryogenic temperatures [44].

### 1.2.3 Heterojunction Barrier Effect

High-injection effects deviate the performance of a bipolar transistor from ideal, because mobile charge concentration is too high for low-injection assumptions to be justified. Among them, the Webster-Rittner effect and the Kirk effect are the most significant high-injection effects for bipolar transistors in general. The former results from the over-simplified assumption of the hole concentration in the base region [45], [46], and the latter is associated with the base push-out at high injection [47]. Both effects give rise to a  $\beta$  roll-off, and the Kirk effect also degrades  $\tau_B$ .

According to a brief calculation in [48], the Webster effect is not pronounced in modern SiGe HBTs, whereas all modern SiGe HBTs are operated in the high-injection regime for the Kirk effect in order to maximize  $f_T$  [48]. In those HBTs where base and collector materials are different, the heterojunction barrier effect (HBE) accompanies the Kirk effect.

As a band edge phenomenon, HBE becomes increasingly important in SiGe HBTs as the operating temperature decreases, due to its thermally-activated nature. HBE was first observed in GaAs/GaAlAs HBTs [49], [50]. Theory predicting the high-injection HBE in SiGe HBTs was presented in [51]. HBE was later investigated in the measurement of SiGe HBTs operating at low temperatures and has been shown to strongly impact both DC and AC device performance [52]. Subsequent research focused on both the optimization of SiGe HBTs in the presence of HBE across temperature [53] and improving compact modeling for circuit designs [54]. These investigations were conducted at temperatures above 77 K, however, and thus HBE was not assessed at very deep cryogenic temperatures such as liquid helium (4.2 K).

To date, the impact of SiGe HBT technology scaling on HBE as a function of temperature has not been carefully addressed. HBE is particularly important for scaling, since both the collector doping and the Ge concentration associated with scaling directly affect HBE. As is demonstrated in this work, an understanding of the

physics of HBE is useful for both compact modeling and device optimization.

### ***1.3 SiGe HBTs for Terahertz Systems***

The semiconductor device community is rapidly entering the terahertz era. Researchers are pushing to demonstrate useful solid-state transistors and the resulting circuits that are capable of operation in the THz regime, defined as sub-millimeter waves from 300 GHz to 3 THz [55].

#### **1.3.1 Terahertz Systems**

The major driving forces for THz solid-state devices are high-frequency communications, radars, and various niche THz applications, as discussed in [55], [56]. In high-frequency communications and radars, higher bandwidth transistors are desirable in a number of applications. For instance, optical fiber communications require active amplifiers in decision circuits, multiplexers, and phase-lock loops operating at 100 GHz clock frequency and above [56]. High  $f_T$  and  $f_{max}$  are also demanded in microwave, millimeter-wave, and submillimeter-wave transceiver designs, where progressive improvements in transistor bandwidth enable the evolution of communications and radar ICs to higher frequency operation. In addition, faster transistors enable the wider-band mixed-signal ICs (e.g., analog-to-digital converters, digital-to-analog converters, etc.) to improve the resolution of radars and communications systems [57]. Meanwhile, THz applications are starting to expand from the initial niche markets of Earth, planetary, and space science [58]-[61] to the larger commercial markets of biomedical imaging, non-metallic object detection, quality control, and secure communications [62]. Of the many THz components currently being used, THz sensors are made from heterodyne semiconductors, heterodyne superconductors, and novel direct detectors (e.g., Schottky diodes and quantum-dot single-photon detectors); THz sources are normally obtained through numerous optical methods [55]. However, lacking available amplifiers and oscillators made of active semiconductor



transistors becomes a bottleneck.

In addition, improved bandwidth in a transistor generally correlates well with improved RF performance (higher gain and lower noise), and thus THz transistors can dramatically widen the design margins of RF through millimeter-wave circuits and systems [63], [64]. A wide variety of space electronics platforms (analog, digital, and RF) designed to operate at space/planetary ambient conditions without bulky and power-hungry heating units can be enabled by high bandwidth transistors.

### 1.3.2 Terahertz Semiconductor Transistors

Recent research has focused on expanding THz options from two-terminal devices (e.g., Schottky diodes) to three-terminal devices (transistors).

Recently, significant progress has been achieved in the frequency response of III-V HBTs and high-electron-mobility transistors (HEMTs). InP-based HEMTs have very recently achieved a peak  $f_T / f_{max}$  of 385 GHz / 1.2 THz [65] and 610 / 305 GHz [66], but practical applications of such devices will be limited by their low breakdown voltages, as compared to InP-based HBTs. The current record for  $f_T$  is 765 GHz for an InP/InGaAs HBT ( $f_{max} = 227$  GHz and  $BV_{CEO} = 1.65$  V (at low injection)) [67], and the record  $f_{max}$  is 755 GHz for another InP-based HBT variant whose  $f_T = 416$  GHz and  $BV_{CEO} = 4.60$  V [68].

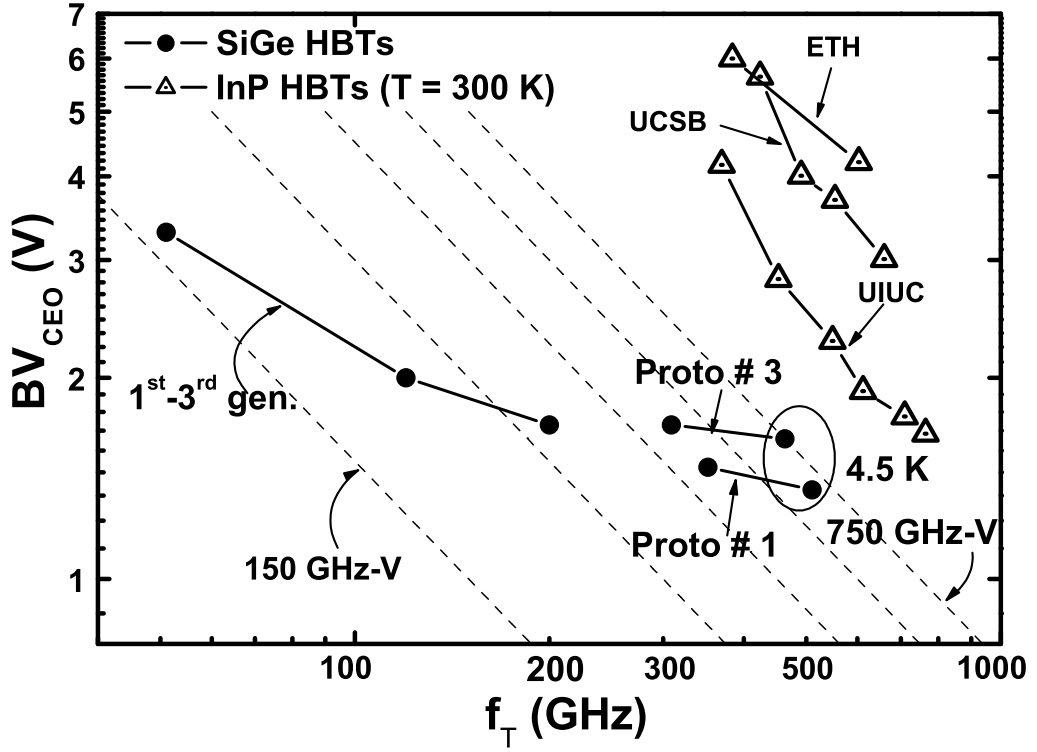
Clearly, it will be highly desirable to achieve these levels of performance in the silicon material system, which will result in the favorable yield, cost, thermal conductivity, integration level, and economy-of-scale associated with silicon IC manufacturing. The higher frequency application of silicon-based transistors is an obvious path enabled by high-performance niche markets first pioneered by III-V transistors. The powerful advantages enjoyed by silicon manufacturing can prove compelling in the end, provided that the silicon-based circuits can satisfy the performance requirements. This trend is evidenced by the successful recent advance of strained Si CMOS

and SiGe HBTs into millimeter-wave radio circuits at 60-100 GHz, the classic domains of III-V devices [69]-[78].

The fastest silicon RF CMOS transistor to-date features a peak  $f_T$  of 360 GHz and  $f_{max}$  of 420 GHz at the 65 nm technology node [79]. Unlike bipolar transistors, for which device-to-device matching generally improves with technology scaling due to the increase in doping levels, CMOS devices face increasingly challenging problems with matching because of the larger relative variations in lateral dimensions at highly-scaled technology nodes. Increased short-channel effects and the high-k dielectrics with metal gate, widely used in the more advanced technology nodes, produce even more serious device matching issues. In addition, the poor output conductance, high leakage currents, degraded low-frequency noise, and low breakdown voltages associated with highly-scaled RF CMOS present serious challenges for circuit designers, often forcing them to explore (time-consuming) circuit innovations to meet circuit and system performance goals.

Meanwhile, fully Si-manufacturing compatible SiGe HBTs are themselves making rapid in-roads into RF through millimeter-wave circuit applications. The possibility of using highly integrated SiGe IC platforms at a much more modest lithography node (SiGe HBTs enjoy about a two-generation lithographic scaling advantage over CMOS for fixed performance) for emerging communications and THz systems is highly appealing from a cost perspective. It is illuminating to examine the performance trends for both SiGe HBTs and III-V HBTs in the  $f_T$ - $BV_{CEO}$  plane (Fig. 3). In this figure, the “UIUC” data were extracted from [67]; the “UCSB” data were provided by Dr. M. Rodwell at the University of California, Santa Barbara; and the “ETH” data were extracted from [85] and [86].

While blind projections from such plots can be misleading, we would point out that trends in SiGe HBT performance suggest that THz levels of performance in SiGe HBTs could be achieved at useful breakdown voltages, while preserving the



**Figure 3:** The  $BV_{CEO}$ - $f_T$  plane of all generations of SiGe HBTs and several examples of record-setting InP-based HBTs, with  $f_T \times BV_{CEO}$  contours indicated.

enormous advantages in integration capability and cost associated with silicon-based manufacturing. It should also be appreciated that the  $BV_{CEO}$  of SiGe HBTs remains roughly constant at all collector current levels, which is significantly different from its III-V HBT counterparts, whose safe operating range of  $V_{CE}$  is almost halved when biased at peak  $f_T$  as compared to low injection.

### 1.3.3 ECL/CML Ring Oscillator Gate Delay

Emitter-coupled logic (ECL) and current mode logic (CML) represent the fundamental building block for modern ultra-high-speed bipolar-based digital systems, and the ECL/CML ring oscillator gate delay ( $\tau_{gate}$ ) remains a simple and powerful metric for assessing overall technology performance, since it provides more information than that which is captured by  $f_T$  and  $f_{max}$  and is hence a better figure-of-merit for digital

circuit applications [39].

The fastest room-temperature  $\tau_{gate}$  reported for Si-based technologies to date is 2.5 ps, with a 300 mV signal swing ( $\Delta V$ ) [80], very close to the best  $\tau_{gate}$  reported for III-V technologies (2.2 ps with a similar  $\Delta V$ ) [81]. The last report of a SiGe ring oscillator being operated below 100 K was over a decade ago, when  $\tau_{gate}$  was found to decrease from 25.4 ps at 300 K down to 21.9 ps at 84 K [82]. Given the superior cryogenic performance of more advanced SiGe HBTs, it is of great interest for researchers to revisit the potential using of SiGe ECL/CML logic for niche cryogenic instrumentation and detection systems [83],[84].

#### 1.3.4 Vertical and Lateral Scaling in SiGe HBTs

The optimized scaling of SiGe HBTs involves a coordinated reduction in both vertical and lateral dimensions. In vertical scaling, one usually decreases the base width ( $W_B$ ) while increasing base and collector doping level ( $N_B$  and  $N_C$ ) to decrease the base, CB SCR, and collector transit times ( $\tau_B$ ,  $\tau_{CSCL}$ , and  $\tau_C$ ). Detailed process innovations to achieve vertical scaling are discussed in [87]. A direct consequence of vertical scaling is the natural increase in  $\beta$ , with peak  $\beta$  increasing monotonically from about 100 in first-generation SiGe HBTs to almost 800 in fourth-generation SiGe HBTs.

The  $f_T$  of an HBT is dependent on vertical scaling, and the contribution of each component to the total  $\tau_{EC}$  has been investigated in [56] and [63]. However,  $\tau_{BE}$ , which represents a trade-off between the transit time through the BE SCR and the charging of the BE junction capacitance, has been neglected by most of the previous works on scaling (e.g. [56], [63], and [38]) partly because it is hard to decouple  $\tau_{BE}$  from  $\tau_B$  and  $\tau_E$  [88]. Only recently did a study show that  $\tau_{BE}$  can be larger than the sum of  $\tau_B$  and  $\tau_C$  [89].

Unfortunately,  $BV_{CEO}$  degrades with increasing  $N_C$ . Also, the vertical scaling

alone inevitably causes larger device parasitics (e.g., base resistance ( $R_B$ ) and CB junction capacitance ( $C_{BC}$ )). While these larger parasitics do not necessarily degrade  $f_T$  at sufficiently high operating currents, they do directly degrade  $f_{max}$ . In a circuit context, a more balanced set of  $f_T$  and  $f_{max}$  is often highly desirable. Therefore, transistors for millimeter-wave amplification need to maintain  $f_{max} > f_T > f_{max}/2$  [90]. To compensate for the  $f_{max}$  degradation that results from vertical scaling, lateral scaling must also be employed to decrease  $R_B$  and  $C_{BC}$ . The higher  $\beta$ ,  $f_T$ , and lower  $R_B$  associated with coordinated device scaling also naturally improve the broadband noise performance, as captured by the minimum noise figure expression in [63]. The base current shot noise and thermal noise caused by  $R_B$  improve with scaling, and the typical room-temperature minimum noise figure ( $NF_{min}$ ) at 15 GHz at 300 K for first, second, third, and fourth-generation SiGe HBTs are 2.6, 1.5, 0.97 and 0.8 dB, respectively.

Shrinking the lateral dimensions, however, presents additional difficulties in  $R_E$  and  $R_C$  reduction because of the smaller junction contact area, and therefore fundamental trade-offs exist and a balanced approach to vertical and lateral scaling must be used. There are ways to mitigate these trade-offs, including raising the extrinsic base and adding a selectively implanted collector (SIC) region [87]. These approaches have been successfully employed to advance the  $f_T$  and  $f_{max}$  (at the same rate of improvement) of commercially-available SiGe HBT technologies.

### 1.3.5 Non-Equilibrium Base Transport

Conventional Si bipolar transistor theory is based on the classic DD equations, and the minority carrier transport in the base is assumed to be dominated by scattering. Therefore the carrier temperature is best reflected in the slope of collector current of the Gummel characteristics at low injection, or approximately,  $\frac{\partial I_C}{\partial V_{BE}} \simeq \frac{qI_C}{kT}$ , or  $\frac{\partial \ln(I_C)}{\partial V_{BE}} \simeq \frac{q}{kT}$ .

With sufficient vertical profile scaling, however, non-equilibrium base transport can occur when the carrier mean free path length ( $l_p$ ) exceeds the neutral base width. This non-classic base transport was predicted in Si BJTs by Monte Carlo simulations [91] and was observed indirectly in AlInAs/InGaAs transistors by fitting  $\beta$  as a function of  $1/W_B$  [92]. The threshold of  $W_B$  for non-equilibrium transport at room temperature is still under investigation, and is predicted to be less than 10 nm [93]. Meanwhile, such non-classical electron transport has been inferred in SiGe HBTs operating at low temperatures [94], where cooling decreases electron-phonon scattering, increases  $l_p$ , and thus makes the base transport more “quasi-ballistic”. In this case, the effective electron temperature, as derived from the slope of  $I_C$ , is expected to increase above the ambient temperature when non-equilibrium transport becomes increasingly important.

#### ***1.4 Negative Differential Resistance***

Research on negative differential resistance (NDR) dates to the advent of semiconductor devices, when different mechanisms of NDR and their applications were studied extensively. Three mechanisms are primarily responsible for NDR in semiconductors, which are: field-excited transfer of electrons from a low-mass valley to a high-mass valley of the conduction band [95]-[97], field induced capture of electrons [98], [99], and Fowler-Nordheim tunneling of electrons in wide-gap materials[100].

These negative resistance effects, however, are intrinsic properties of the related III-V materials, and thus are irrelevant to the group-IV semiconductors (e.g., Si and Ge). Meanwhile, semiconductor tunnel structures that exhibit NDR can be found in both III-V and group-IV devices, examples including Esaki diodes [101], double-barrier resonant tunneling diodes [102], single-barrier tunneling diodes [103], [104], and resonant interband tunneling devices [105]. The integration of resonant tunneling devices into commercially-available IC processing (e.g., in Si) has been

recently pursued for mixed-signal and fast digital (logic/memory) applications [106]-[113]. Other silicon-based NDR transistors include such devices as bistable gated bipolar transistors that are of potential interest to digital circuits [114], [115]. Compared to all such NDR devices, the NDR effects described in the current work have a fundamentally different physical origin.

## **1.5 Higher Breakdown Voltage Design**

In circuit applications, a higher transistor breakdown voltage enables higher output power and power added efficiency in power amplifier, as well as better signal-to-noise ratio and more circuit topology options in all circuit designs. For bipolar transistors, three types of breakdown voltages are particularly important: Open-base breakdown voltage ( $BV_{CEO}$ ) represents the worst case scenario, and  $BV_{CEO}$  is usually recommended by design kits as the maximum rail voltage ( $V_{CC}$ ); CB junction breakdown voltage ( $BV_{CBO}$  or  $BV_{CBS}$ ) represents the best case scenario, and even the best circuit topology cannot be biased higher than  $BV_{CBO}$ ; the critical voltage ( $BV_{CB,crit}$ ) happens when current crowding reaches pinch-in.  $BV_{CB,crit}$  is always between  $BV_{CEO}$  and  $BV_{CBO}$ , and is usually closer to  $BV_{CBO}$ . While  $BV_{CBO}$  is only related to the multiplication factor in the CB junction,  $BV_{CEO}$  is also determined by  $\beta$ . The modeling of  $BV_{CB,crit}$  is more complicated, and involves 3-D analysis. However, lowering impact ionization in the CB junction improves all three breakdown voltages.

### **1.5.1 Speed / Breakdown Voltage Trade-Off**

In device scaling, there has always been an increasing interest in pursuing higher breakdown voltages in semiconductor devices. For example, a well-documented trade-off exists between the device on-state resistance and the breakdown voltage in high-power device designs.

In our high-speed devices, there also exists a fundamental trade-off between the device speed (e.g.  $f_T$ ) and breakdown voltage (e.g.  $BV_{CEO}$  and  $BV_{CBO}$ ) [118], [119].

Normally higher  $f_T$  can be achieved by increasing the doping level of the SIC region. In this way, the CB SCR region shrinks, giving rise to shorter carrier transit times. However, such a method degrades the impact ionization within the CB SCR, causing higher multiplication factors (M-1) and lower breakdown voltages ( $BV_{CEO}$  and  $BV_{CBO}$ ), which can potentially compromise both the RF output power and the signal-to-noise ratio of the relevant systems. This issue has already become critical in highly-scaled SiGe HBTs, whose  $f_T$  and  $f_{max}$  have well exceeded the 100 GHz regime. Recent research bodes well for further scaling toward half-terahertz (500+ GHz) performance at room temperature [80]. In pursuit of higher  $BV_{CEO}$  with little degradation in the AC performance, device engineers have had to sacrifice half of  $\beta$  for about 0.1 V increase in  $BV_{CEO}$ .

More complicated collector profile designs have been proposed in GaAs HBTs to improve the speed / breakdown voltage trade-off. For example, [120] uses a i-p<sup>+</sup>-n<sup>+</sup> collector to form an inverted field that keeps the electrons staying in the  $\Gamma$ -valley to improve the device speed. Nevertheless, this method is for GaAs devices only, and does not apply to Si BJTs and SiGe HBTs. The work in [121] uses a double-collector design to form a high-low doping profile in the collector of an AlGaAs/GaAs HBT. This method has been tried many times in state-of-the-art SiGe HBTs and has been proven ineffective through numerous simulation efforts.

### 1.5.2 History of Superjunctions

The superjunction is formed with multiple alternating p-type and n-type thin layers. The idea of using a superjunction to improve the breakdown voltage has been applied to semiconductor power devices for two decades [122]-[128].

In a conventional PN junction, the electric field peaks at one point only, where impact ionization degrades significantly and causes breakdown to happen. By using alternating doping types in the high-electric field region, the superjunction flattens



out the electric field inside the SCR and therefore sustains a higher voltage across the SCR at the same impact ionization rate. This design facilitates the use of high-power devices with higher doping level in the substrate, which results in lower on-state resistivity. In fact, all the designs in [122]-[128] use the superjunction in the body or substrate of a field effect transistor, a bipolar transistor, a PIN diode, or a rectifier to improve the trade-off between the on-state resistivity and the off-state breakdown voltage.

For bipolar transistors, a superjunction was used in the depletable lateral collector region of a bipolar transistor to improve the trade-off between the collector resistance and the breakdown voltage [129]. This design is for high power bipolar transistors only, as can be clearly seen from the lateral depletable collector structure. Such a design does not improve device speed and RF or millimeter-wave frequency response at a fixed breakdown voltage.

All the superjunction devices mentioned above are high-power devices, in which very high bias voltages (above 50 V) are usually desired, and device performance above 1 GHz is not a concern. The superjunction is used to improve the trade-off between the device off-state breakdown voltage and the on-state resistance.

## ***1.6 From Drift-Diffusion to Monte Carlo***

The semiclassical particle (an electron or a hole) in semiconductor devices is an approximation of quantum mechanical waves. The equations that are used to describe the carrier behavior in device physics and TCAD simulations represent different levels of approximations, each sacrificing calculation accuracy for acceptable simulation complexity.

The most fundamental equations that govern the trajectory of an electron in the phase space (position and momentum space) are listed below [130]-[132]:

$$\frac{d\vec{r}}{dt} = \frac{1}{\hbar} \nabla_{\vec{k}} E(\vec{k}), \quad (1)$$

$$\frac{d\hbar\vec{k}}{dt} = -\nabla_r E_C(\vec{r}, t) + F_S(\vec{r}, t). \quad (2)$$

Equation (1) describes the velocity of the semiclassical particle in the position space, where  $\vec{k}$  is the crystal momentum, and  $E(\vec{k})$  represents the band structure of the semiconductor. Equation (2) looks like Newton's law in the momentum space, where  $E_C$  is bottom of the conduction band, and  $F_S(\vec{r}, t)$  represents the random force caused by various carrier scattering mechanisms. By using a distribution function,  $f(\vec{r}, \vec{k}, t)$ , one reaches the Boltzmann transport equation (BTE):

$$\frac{\partial f}{\partial t} + \vec{v} \cdot \nabla_r f - \frac{q\vec{E}}{\hbar} \nabla_k f = \hat{C}f. \quad (3)$$

where  $\vec{E}$  is the electric field and  $\hat{C}$  the scattering term. The zeroth moment of  $f(\vec{r}, \vec{k}, t)$  leads to the continuity equation for electron density:

$$\frac{\partial n(\vec{r}, t)}{\partial t} = \frac{1}{q} \nabla_r \cdot \vec{J}_n + G_n - R_n, \quad (4)$$

where,  $J_n = -nq\vec{v}$ ;  $G_n$ ,  $R_n$ , and  $\vec{v}$  are current density, electron generation rate, electron recombination rate, and the average velocity, respectively. With proper assumptions [130]-[132] and assuming  $\mu$ ,  $D_n$ , and  $\vec{E}$  are the electron mobility, the diffusion coefficient, and electric field, respectively, the first moment leads to the drift-diffusion (DD) equation:

$$\vec{J}_n = nq\mu\vec{E} + qD_n\nabla_r n. \quad (5)$$

Equation (5) assumes that carriers are in thermal equilibrium, and that variables such as field-dependent mobilities and diffusivities respond instantaneously to changes in the electric field. If non-local effects such as velocity overshoot are required to be considered, equation (5) must be modified to include carrier temperatures that are different from the lattice temperature [133].

The second and third moments of the distribution function give the continuity equations for the energy density and the energy flux:

$$\frac{\partial nw}{\partial t} + \nabla_r \cdot \vec{J}_W - qn\vec{v} \cdot \vec{E} = C_W; \quad (6)$$

$$\vec{J}_W = nw\mu_E\vec{E} + \nabla_r (nD_E w). \quad (7)$$

where  $C_W$  is the energy collision term;  $w$  is the average carrier energy;  $\vec{J}_W$ ,  $\mu_E$  and  $D_E$  are carrier energy flux, energy transport mobility and energy diffusion coefficient respectively. Details of equations (6) and (7) can be found in [130]. In conventional SiGe HBTs at room temperature, where the electron kinetic energy is assumed to respond instantaneously to the magnitude of the electric field, equations (4), (5) and Poisson's equation are sufficient for describing the carrier behavior. In fact, DD has been used to explain the physics of most bipolar semiconductor devices in the past few decades [134].

Under high electric fields or in highly-scaled transistors, however, the electron temperature can be higher than the lattice temperature. Therefore, equation (5) needs to be modified to account for non-local effects in the CB junction, as well as for

the base transport. In addition, higher moments of BTE (equations (6) and (7)) are required for hydrodynamic simulation. When both DD and hydrodynamic approaches fail to be accurate enough in the aggressively-scaled SiGe HBTs, one needs to return to equations (1) and (2) by averaging the results from a large number of simulated trajectories. This method is called the Monte Carlo technique [131], [135].

## CHAPTER II

### NOVEL NEGATIVE DIFFERENTIAL RESISTANCE EFFECT AND COLLECTOR-CURRENT KINK EFFECT

Since the NASA ETDP project employs IBM 5AM (similar to 5HP described in Chapter 1) SiGe HBT BiCMOS technology, all data presented in this chapter are from the commercially-available 5AM unless specifically indicated. The technology details of the first, third, and fourth-generation SiGe HBTs can be found in [10], [12], and [13]. The AC and DC test structures were specially designed for NASA ETDP, and information regarding test structure designs can be found in [30].

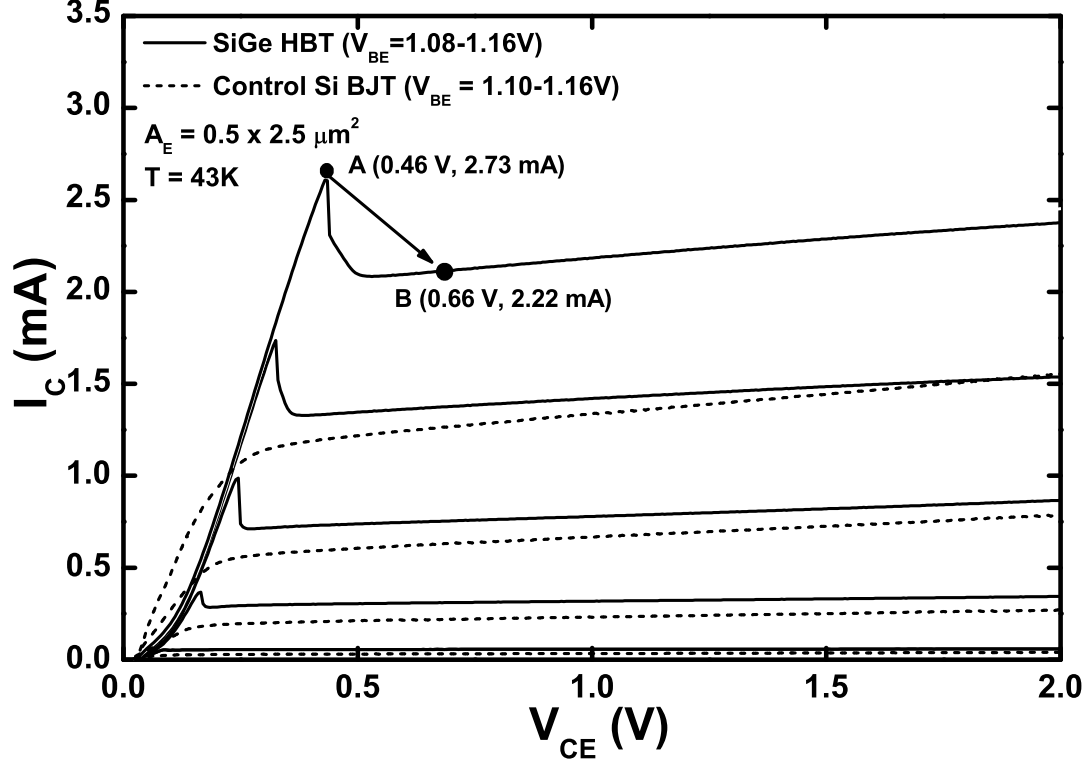
#### *2.1 Novel Experimental Effects*

The cryogenic device measurements were made using an on-wafer, open-cycle, liquid helium cryogenic probe system capable of DC to 40 GHz operation from 350 K to 5 K. The system thermometry was calibrated by comparing device characteristics inside the system with those measured directly immersed in both liquid nitrogen (77.3 K) and liquid helium (4.2 K). The temperature accuracy is believed to be better than 1 K, and is stable at intermediate temperatures. An Agilent 4156 Semiconductor Parameter Analyser was used for DC device characterization, and an Agilent 8510C Vector Network Analyzer was used for AC measurements. Standard calibration and de-embedding techniques were used at each measurement temperature.

##### **2.1.1 Negative Differential Resistance Effect**

A new device effect, a negative differential resistance (NDR) region in the forced- $V_{BE}$  output characteristics, is consistently observed of these first-generation SiGe HBTs operating at high-injection at deep cryogenic temperatures (Fig. 4). A control Si BJT

(with nearly the same doping profile but without Ge) was also measured and no NDR is observed in Fig. 4, indicating that this NDR effect is clearly Ge-induced.

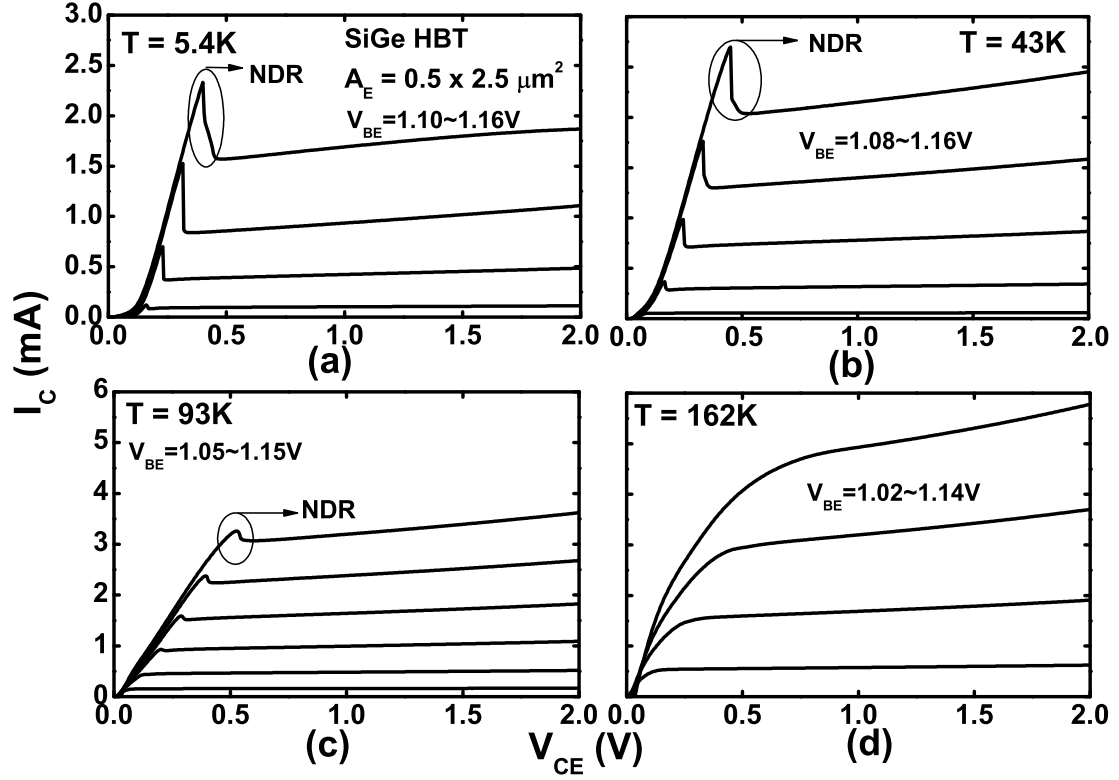


**Figure 4:** Comparison of the measured forced- $V_{BE}$  output characteristics of the SiGe HBT and the control Si BJT at 43 K.

The magnitude of NDR in a first-generation SiGe HBT, is obviously temperature dependent, increasing in magnitude as temperature decreases from 93 K to 5.4 K (Fig. 5). The NDR is almost negligible at 93 K, and disappears entirely at temperature  $> 93$  K. The excess collector current ( $\Delta I_C$ ) of the SiGe HBT is defined as the collector current difference directly before and after the observed NDR occurs, or,

$$\Delta I_C = I_{C,peak} - I_{C,valley}, \quad (8)$$

where  $I_{C,peak}$  (the peak current) is the maximum collector current immediately before



**Figure 5:** Measured forced- $V_{BE}$  output characteristics of a SiGe HBT at (a) 5.4 K, (b) 43 K, (c) 93 K, and (d) 162 K. NDR is observed in (a)–(c).

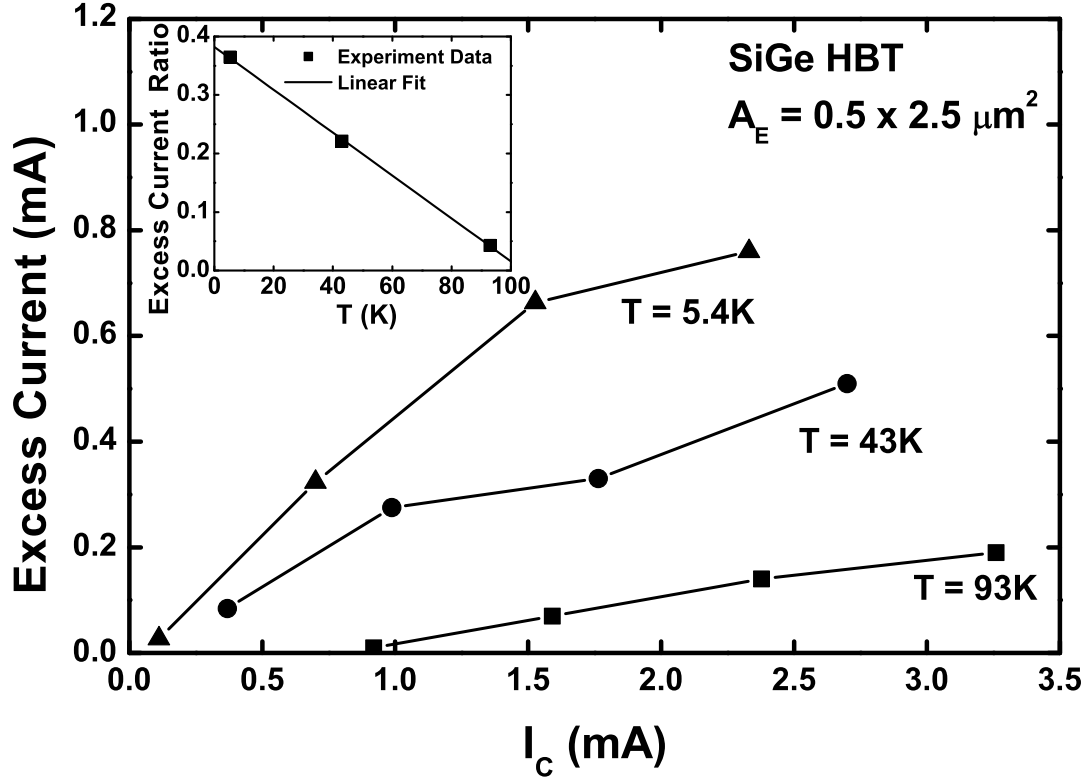
the NDR, and  $I_{C, valley}$  (the valley current) is the minimum collector current immediately after the NDR. The  $\Delta I_C$  is proportional to  $I_{C, peak}$  to the first order (Fig. 6). To quantitatively assess the temperature dependence of the NDR, an excess current ratio ( $ECR$ ) is defined as the average of all the measured  $\Delta I_C$  normalized by  $I_{C, peak}$ , or,

$$ECR = Avg \left( \frac{\Delta I_C}{I_{C, peak}} \right) = Avg \left( \frac{I_{C, peak} - I_{C, valley}}{I_{C, peak}} \right) = Avg \left( 1 - \frac{I_{C, valley}}{I_{C, peak}} \right), \quad (9)$$

where  $Avg()$  means the mathematical average value.

In the literature, peak-to-valley current ratio ( $PVCR$ ) is more often used as a figure-of-merit for NDR devices, and is usually defined as,

$$PVCR = \frac{I_{C, peak}}{I_{C, valley}}, \quad (10)$$



**Figure 6:** The excess current for NDR (defined as the  $I_C$  difference before and after the NDR occurs) in the forced- $V_{BE}$  output characteristics as a function of collector current at 5.4 K, 43 K and 93 K. Inset shows the linear fit for the excess current ratio (defined as the excess current divided by  $I_C$  before the NDR) across temperatures.

For convenience and better insight,  $ECR$  is used in the present thesis, but can be easily converted to  $PVCR$  using,

$$ECR = Avg \left( 1 - \frac{1}{PVCR} \right), \quad (11)$$

The extracted  $ECR$  linearly increases as temperature decreases from 93 K to 5.4 K, as shown in the inset of Fig. 6. The  $ECR$  as a function of temperature ( $T$ ) can be quantitatively expressed as,

$$ECR(T) = a \cdot T + b, \quad (12)$$

with fitting parameters  $a = -0.00367 \text{ K}^{-1}$  and  $b = 0.38282$ . Equation (12) can be used to predict the magnitude of NDR for any given collector-current level at



any temperature for this SiGe technology, and can be similarly applied to any SiGe process technology as a design aid. For reference, the resultant  $PVCR$  is 1.3 at 43 K and rises to 1.6 at 5.4 K. These are reasonable values for NDR, given that similar  $PVCR$  numbers are reported in resonant tunneling diodes (Table 2).

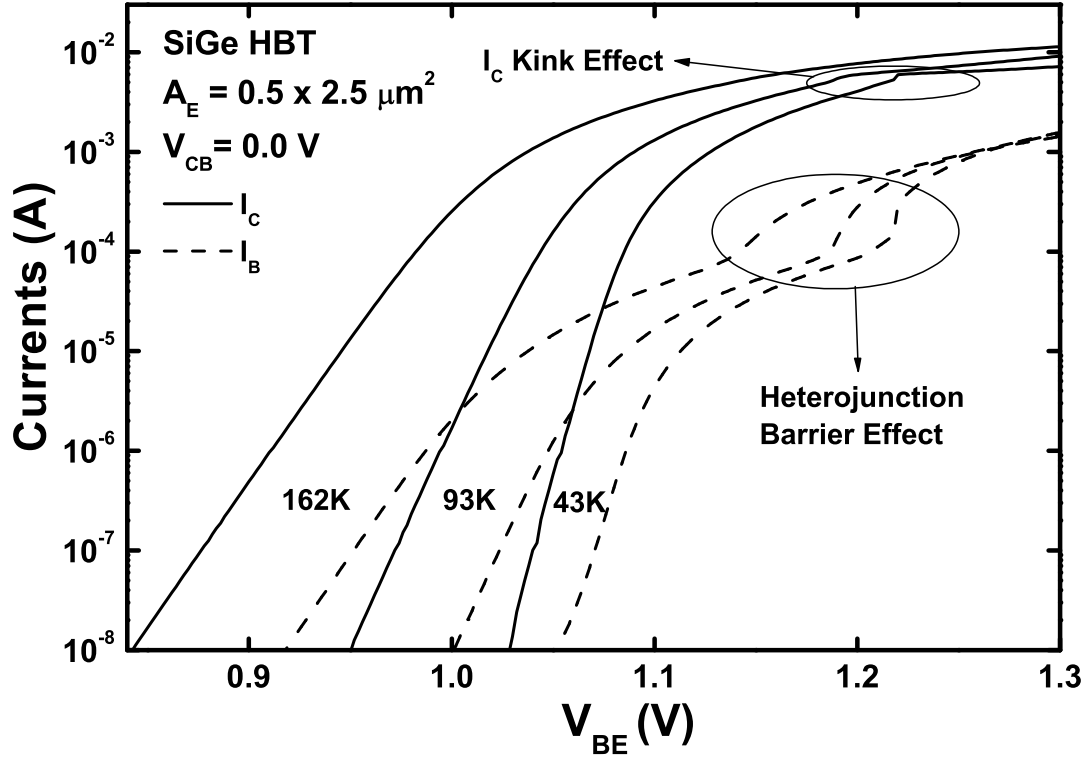
**Table 2:** Comparison of  $PVCR$  in SiGe HBTs and in different Si resonant tunneling diodes

Reference	Device	PVCR	Temperature (K)
This work	SiGe HBT	1.3	43
		1.6	5.4
[116]	Si/SiGe double barrier diodes	1.2	room-T
		1.5	77
[106]	Si/SiGe RTD	2.9	room-T
[117]	Si Esaki diode	1.12 - 1.08	4.2 - 325

### 2.1.2 Collector-Current Kink Effect

An anomalous collector current kink is also observed in the Gummel characteristics of the SiGe HBT at temperatures below about 100 K (Fig. 7). The  $I_C$  kink effect is characterized by a sharp  $I_C$  increase at a given  $V_{BE}$ , resulting in a large “spike” in the transconductance ( $g_m$ ), as shown in the dotted curve in Fig. 8).

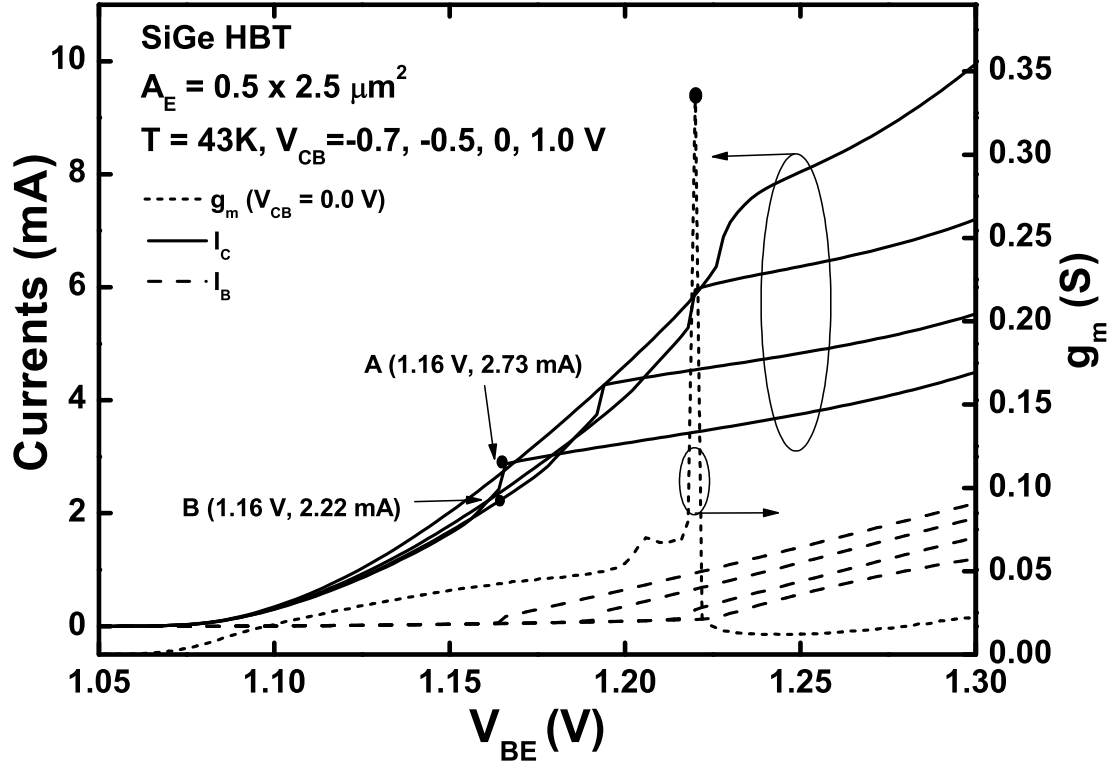
From a careful examination of the temperature-dependent Gummel characteristics in Fig. 7, the  $I_C$  kink can be seen to be clearly associated with the onset of classical HBE, since the  $I_C$  kink occurs at the identical  $V_{BE}$  as the sharp  $I_B$  increase produced by HBE. However, the observed  $I_C$  increase is fundamentally different from the impact of conventional HBE on the device, where  $I_C$  is clamped (decreases) because of the induced conduction band barrier at the SiGe-Si base-collector heterojunction [52], [54]. Note that conventional HBE theory predicts a resultant decrease, not an increase, in  $I_C$  under these conditions, and thus is clearly different from what have been previously reported in the context of HBE.



**Figure 7:** Measured Gummel characteristics of the SiGe HBT at 43 K, 93 K, and 162 K. The  $I_C$  kink effect can be observed at 93 K and below, together with the sudden  $I_B$  increase representing the onset of heterojunction barrier effect (HBE).

Neither the  $I_C$  kink nor the resultant  $g_m$  spike can be found in the control Si BJT at 43 K (Fig. 9), again highlighting the Ge-dependence of this unique effect. In addition, measurements of the emitter-base diode of the SiGe HBT over temperature, while leaving collector terminal open, shows no such kink effect in the diode current, which further demonstrates that the effect is not associated with the emitter-base (EB) junction.

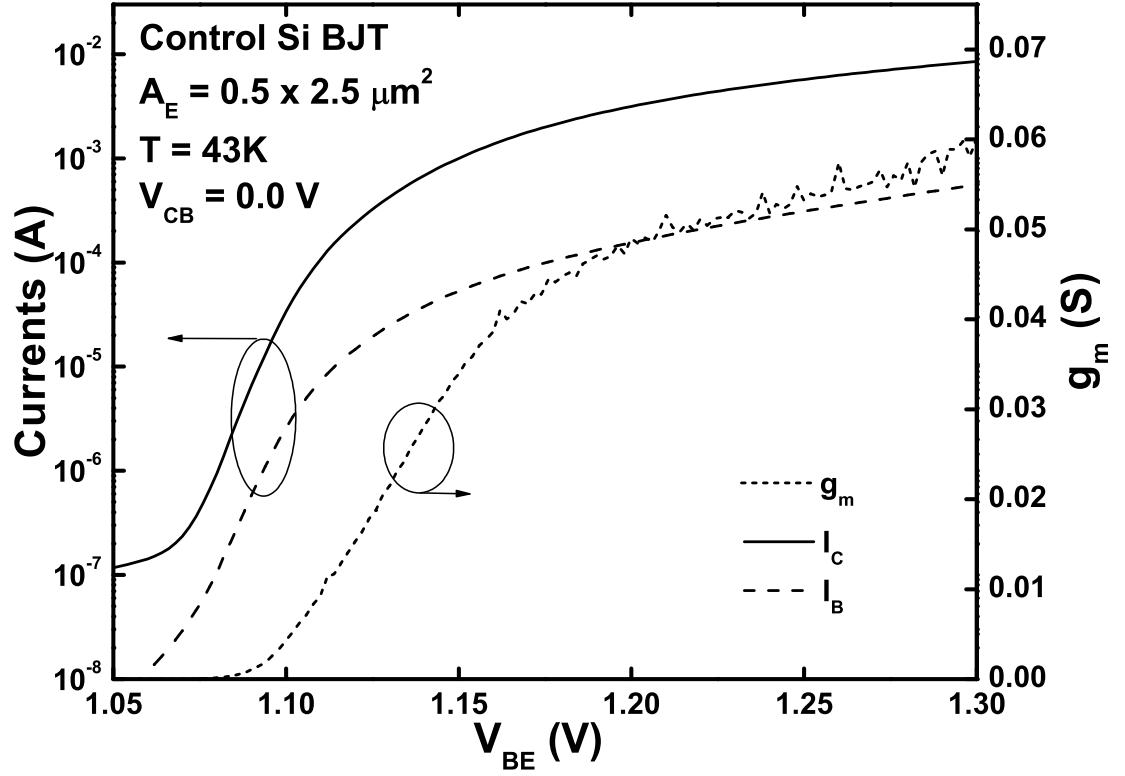
In fact, the NDR observed in the forced- $V_{BE}$  output characteristics, and the  $I_C$  kink in the Gummel characteristics, are produced by the same physical mechanism. Fig. 8 shows the Gummel characteristics at different  $V_{CB}$ 's on a linear scale, and indicates that a smaller  $V_{CB}$  results in a suddenly (and unexpectedly) higher  $I_C$  at certain bias points (e.g., points A and B in Figs. 4 and 8), since the sudden increase in



**Figure 8:** Measured Gummel characteristics (with currents on a linear scale) of the SiGe HBT at 43 K and at different  $V_{CB}$ 's. Also shown is the transistor transconductance ( $g_m$ ) at  $V_{CB} = 0$  V. Points A and B correspond to points A and B in Fig. 4, indicating the intrinsic relationship between NDR and the kink effect.

$I_C$  occurs at smaller  $V_{BE}$  for lower  $V_{CB}$ . This produces an NDR region in the forced- $V_{BE}$  output characteristics, where higher  $V_{CE}$  produces a lower  $I_C$ , as illustrated explicitly in Fig. 4.

Finally, any possibility of device oscillation during measurement as a physical origin of the observed NDR has been systematically eliminated. The fact that a control Si BJT does not show these novel NDR effects is important in this context, because the Si BJT is identical to the SiGe HBT in every way including the device and pad layout, and used an identical measurement setup. Various additional techniques were used to ensure the robustness of the data. First, measurements were repeated on several different experimental setups, including: 1) an on-wafer cryogenic



**Figure 9:** Measured Gummel characteristics of the control Si BJT at 43 K and at  $V_{CB} = 0\text{ V}$  and  $V_{CB} = -0.5\text{ V}$ . Also shown is the transistor transconductance ( $g_m$ ) at  $V_{CB} = 0\text{ V}$ .

probe station with measurements using co-linear DC probes that are designed specifically for low temperature measurements; 2) an on-wafer cryogenic probe station with measurements using G-S-G shielded probes on different AC test structures; and 3) a closed-cycle helium cryostat system using 28-pin DIP packages and wire-bonded samples. These three distinct setups produce consistent and repeatable results. Second, test structures with different device layouts were measured. Besides the standard test structures provided by IBM, different AC and DC test structures are also designed and measured, and all show identical results. Third, devices of different geometries were measured. Large-sized HBTs are known to be generally more immune to oscillations because of the smaller impact of layout parasitics. In addition to the standard

0.5×2.5 and 0.5×1.0  $\mu m^2$  SiGe HBTs, we also measured the 0.5×20×2  $\mu m^2$  power transistor. All of these devices show the same novel NDR effects. Fourth, different measurement techniques were used in the 4155/4156 Semiconductor Parameter Analyzer in order to verify that the data are robust. For instance, sweeping  $V_{BE}$  in both directions produces exactly the same Gummel characteristics, and sweeping  $V_{CE}$  in both directions produces identical output characteristics. “Long” *vs.* “short” integration times on the 4155/4156 was also chosen for the NDR and  $I_C$  kink effect regions of the curves, showing no differences. Measuring a single bias point over a certain period of time was also adopted in the NDR and IC kink effect regions, and produced stable and reproducible results. In summary, we have taken great pains to ensure that the reported data are robust, and we have high confidence that the observed effects are in fact real.

## 2.2 *Review of the Heterojunction Barrier Effect*

The measurement results in the previous section strongly suggest a direct connection between the observed NDR and  $I_C$  kink with classical HBE. However, since conventional HBE theory in HBTs predicts only an  $I_C$  decrease and a sudden increase in  $I_B$  at a  $V_{BE}$  ( $J_C$ ) sufficient to turn-on HBE for any fixed  $V_{CB}$  [54], we must revisit the physical origins of HBE, and explicitly examine how device operation at deep cryogenic temperatures can change the coupling of HBE to the transistor response.

In homojunction bipolar transistors, high-injection performance is mainly limited by Kirk effect [47] and Webster-Rittner effect [45], [46]. In heterojunction bipolar transistors operating at high injection, however, the combination of Kirk effect and HBE limit device performance. Physically, in HBE, when  $J_C$  gets sufficiently large, the increased minority carrier concentration in the CB space charge region acts to compensate the local ionized donors, causing the original CB electric field to collapse. Classical Kirk effect pushes the boundary of the CB space charge region further into

the selectively implanted collector (SIC) region, gradually exposing the Si/SiGe heterojunction interface that was originally masked by the band bending in the CB space charge region. The exposure of the valence band barrier induces a conduction band barrier that blocks the hole injection into SIC needed to maintain charge neutrality, producing a pile-up of holes at the hetero-interface. This local hole accumulation in-turn leads to a conduction band barrier that opposes electron transport from the base to the collector, rapidly degrading both DC and AC device performance.

Based on the quantitative theory of *Liang et al.* [54], the deviation of charge in the base ( $\Delta Q_b$ ) as a function of  $J_C$  due to the barrier effect can be expressed as follows,

$$\Delta Q(J_C) = C \cdot J_{C,barrier} \cdot (e^{\Delta E_{C,barrier}/2kT}) \cdot \left[ \frac{J_C}{J_{C,barrier}} - \ln\left(\frac{J_C}{J_{C,barrier}}\right) - 1 \right] \quad (13)$$

where  $C$  is a (constant) fitting parameter,  $\Delta E_{C,barrier}$  is the height of the conduction band barrier, and  $J_{C,barrier}$  is the critical onset current density for barrier effect. Note that  $J_{C,barrier}$  is a monotonically increasing function of  $V_{CB}$ , since  $\Delta E_{C,barrier}$  gradually decreases as  $V_{CB}$  increases. This is because the conduction band edge ( $E_C$ ) is effectively “pulled down” by the higher  $V_{CB}$  imposed on the space charge region of the CB junction via Poisson’s equation and the continuity equations.

Consequently, this extra base charge storage induces not only a rapid current gain ( $\beta$ ) roll-off (degradation) but also a sudden decrease in  $f_T$  and  $f_{max}$  [39]. More fundamentally, the high-injection base transport under classical HBE differs from that of Si BJT in the following ways: the electron diffusion current component decreases (sometimes even to negative value) due to a reduction in the electron density gradient in the base, as shown in Fig. 10, and the drift current component now dominates  $I_C$  because of the large carrier density. In addition, the extra HBE-induced accumulated electron/hole charges act as a local dipole, imposing an inverse electric field that acts to weaken the drift current in this region.  $I_C$  is thus strongly reduced under HBE, while  $I_B$  rises rapidly due to enhanced recombination, strongly degrading the current gain and  $f_T$  of the device.

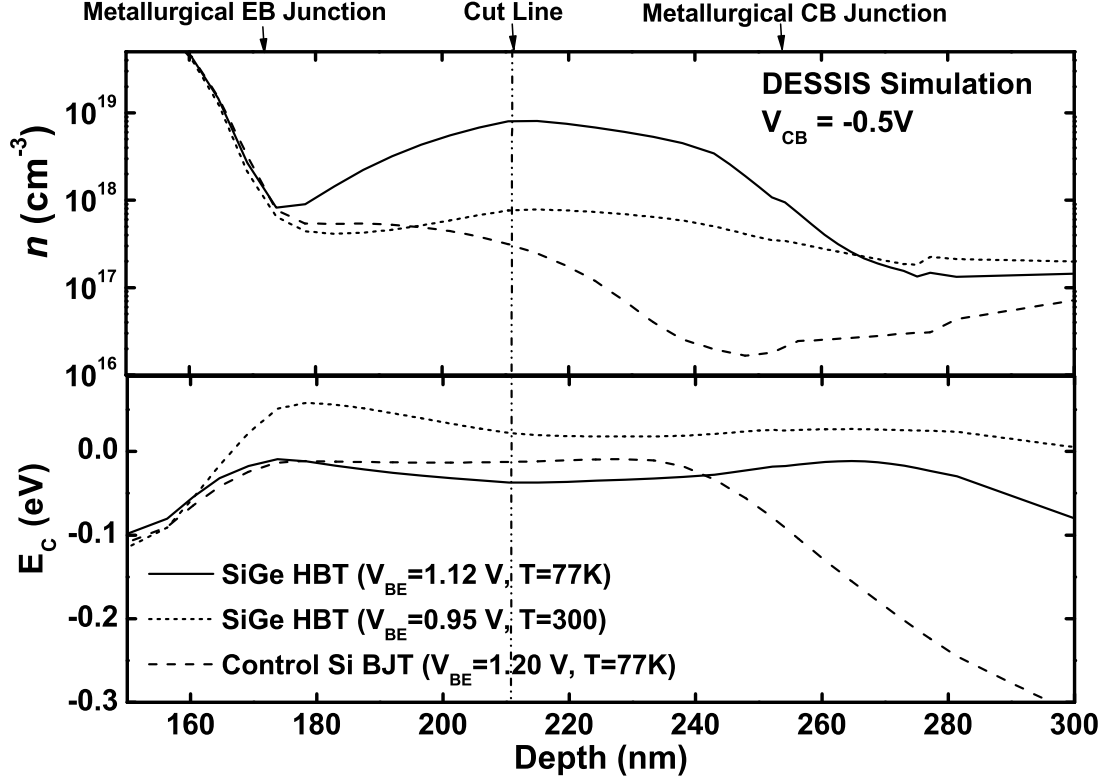
### 2.3 *Enhanced Positive Feedback at Deep Cryogenic Temperatures*

HBE at deep cryogenic temperatures is more complicated. The impact of the heterojunction barrier on the charge accumulation is generally larger at very low temperatures because HBE is a thermally-activated band-edge phenomenon, leading to much higher electron concentrations in the base at low temperatures than at higher temperatures (at say fixed  $J_C$ ). A comparison of the simulated conduction band edge ( $E_C$ ) and electron density ( $n$ ) are given in Fig. 10, where about 10 times more HBE-induced electrons are accumulated in the base at 77 K than at 300 K at a similar current density. In the base region of the SiGe HBT, the electron current density  $J_n$  can be expressed as,

$$\vec{J}_n = q \cdot \mu_n \cdot \nabla \left( \frac{n \cdot k \cdot T}{q} \right) + q \cdot n \cdot \mu_n \cdot \vec{E} \quad (14)$$

where the first term of  $\vec{J}_n$  represents the diffusion component of the total electron current density, and the second term is the drift current component. Therefore, significantly larger drift current component is expected at cryogenic temperatures.

With the increase of the collector current density at higher temperatures, the inverse electric field caused by the conduction band barrier and Kirk effect mutually opposes the further increase of the electrons flowing from emitter to collector; consequently the barrier height, the electron concentration, and the electron current density form a negative feedback loop that acts to “stabilize” the total collector current. At deep cryogenic temperatures, however, the stronger HBE-induced electron concentration in the base adds up to the total drift current overcoming the negative influence of the inverse field on  $J_C$ , which in turn greatly enhances the heterojunction barrier as well as its impact on the collector current density. As a result, the barrier height, the electron concentration, and the electron current density act to produce a positive feedback loop at the onset of HBE. This greatly accelerates the formation of the heterojunction barrier, yielding a much more rapid electron accumulation at the



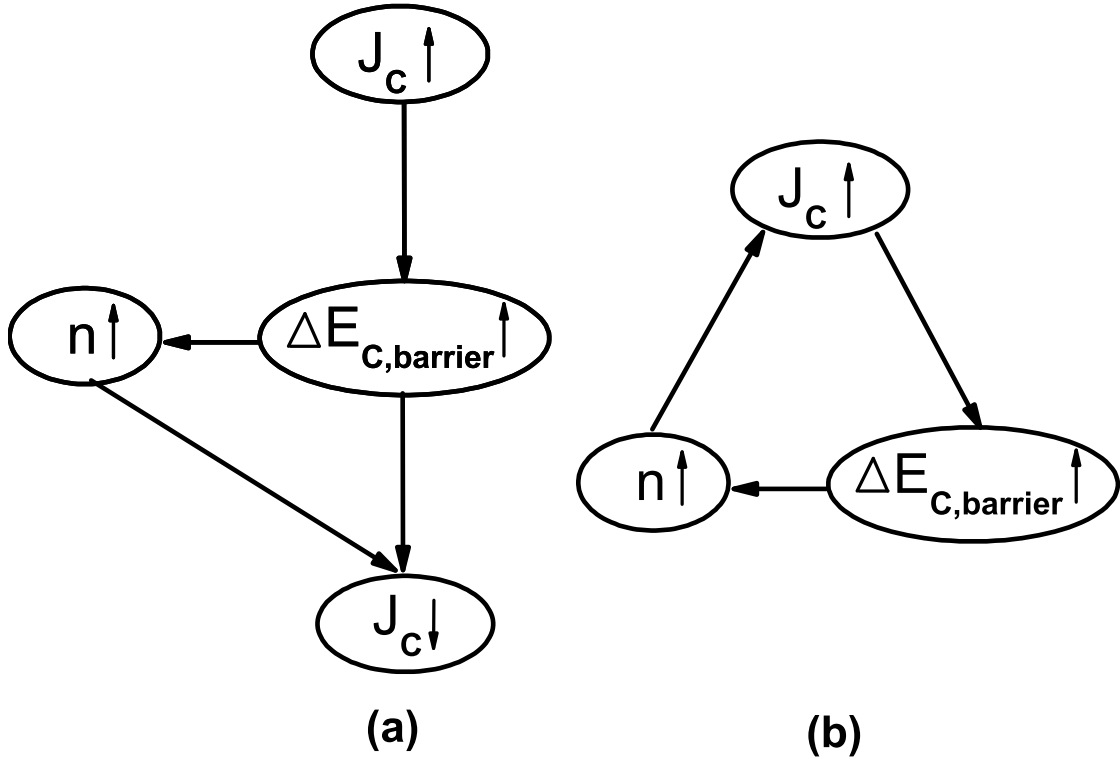
**Figure 10:** Simulated conduction band edge ( $E_C$ ) and electron density ( $n$ ) as functions of depth for the SiGe HBT at 77 K, at 300 K and the control Si BJT at 77 K. The  $V_{BE}$ 's are chosen to ensure comparable collector current for all the three situations. Positions for the metallurgical junctions and the cut line are marked.

onset of the HBE, which not only decreases the diffusion current component, but also enhances the drift current component at a much higher rate, eventually producing a sudden  $I_C$  increase in the Gummel plot for a fixed  $V_{CB}$ . The transformation of the negative feedback at higher temperatures into the positive feedback at deep cryogenic temperatures is illustrated in Fig. 11 (a) and (b). This induced positive feedback loop gets increasingly stronger (enhanced) as the temperature further drops.

The observed NDR behavior in these SiGe HBTs is a direct consequence of the above process. During the forced- $V_{BE}$  measurement, the conduction band barrier gradually decreases as  $V_{CE}$  (and thus  $V_{CB}$ ) increases, because ( $E_C$ ) is effectively “pulled down” by the higher  $V_{CB}$  imposed on the space charge region of the CB



junction. As a result, the number of the accumulated electrons in the base region decreases. The barrier height is more sensitive to  $V_{CB}$  at lower temperatures also due to the enhanced positive feedback between the barrier height, the charge accumulation, and the current density with cooling, as shown in Fig. 11(b). This causes a more rapid decrease of  $n$  when the HBE is turned off by increasing  $V_{CE}$  at low temperature, giving rise to an NDR in the output characteristics.

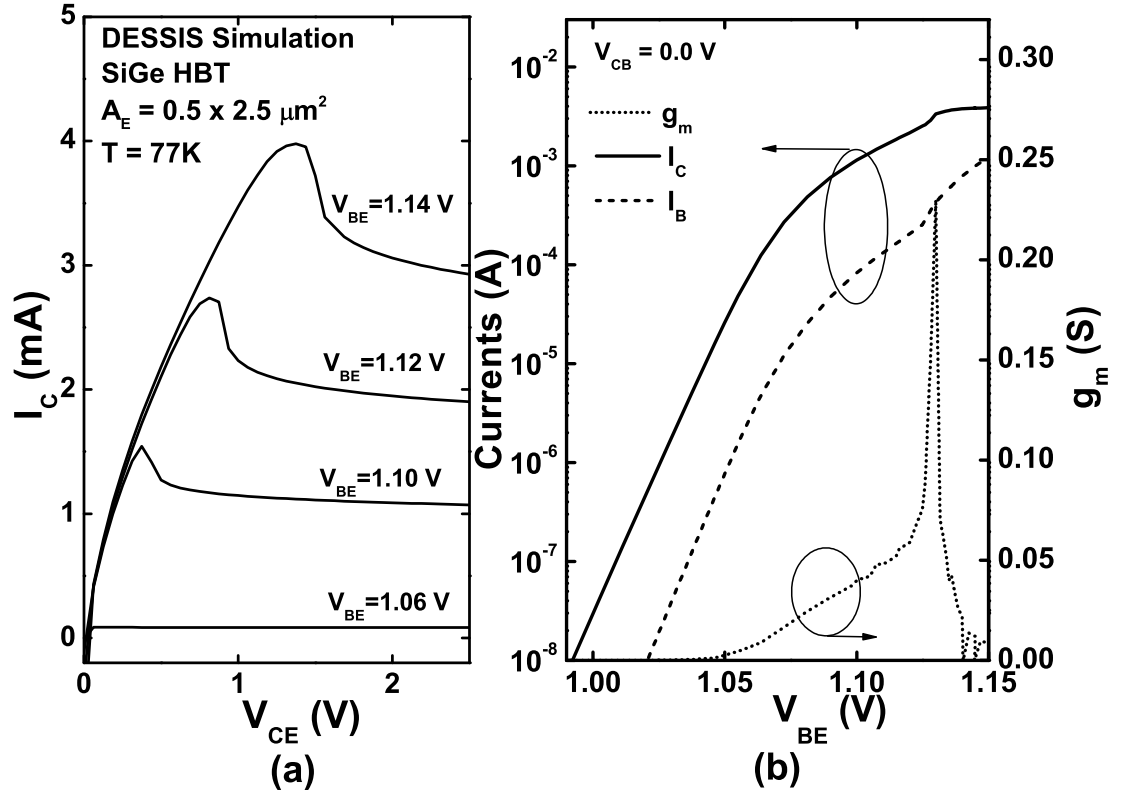


**Figure 11:** The feedback loop of the key physical parameters happening at heterojunction barrier effect and at (a) higher temperatures forming a negative feedback, (b) deep cryogenic temperatures forming a positive feedback.

In addition, cooling effectively turns-on HBE at a higher  $J_C$  than at 300 K due to the shift in the built-in potential of the CB junction [39], together with the slight increase in the electron saturation velocity at low temperature [5]. These effects further contribute to the total enhancement of the positive feedback of the loop depicted in Fig. 11(b) at cryogenic temperatures, effectively magnifying the phenomena.

## 2.4 TCAD Simulation Results

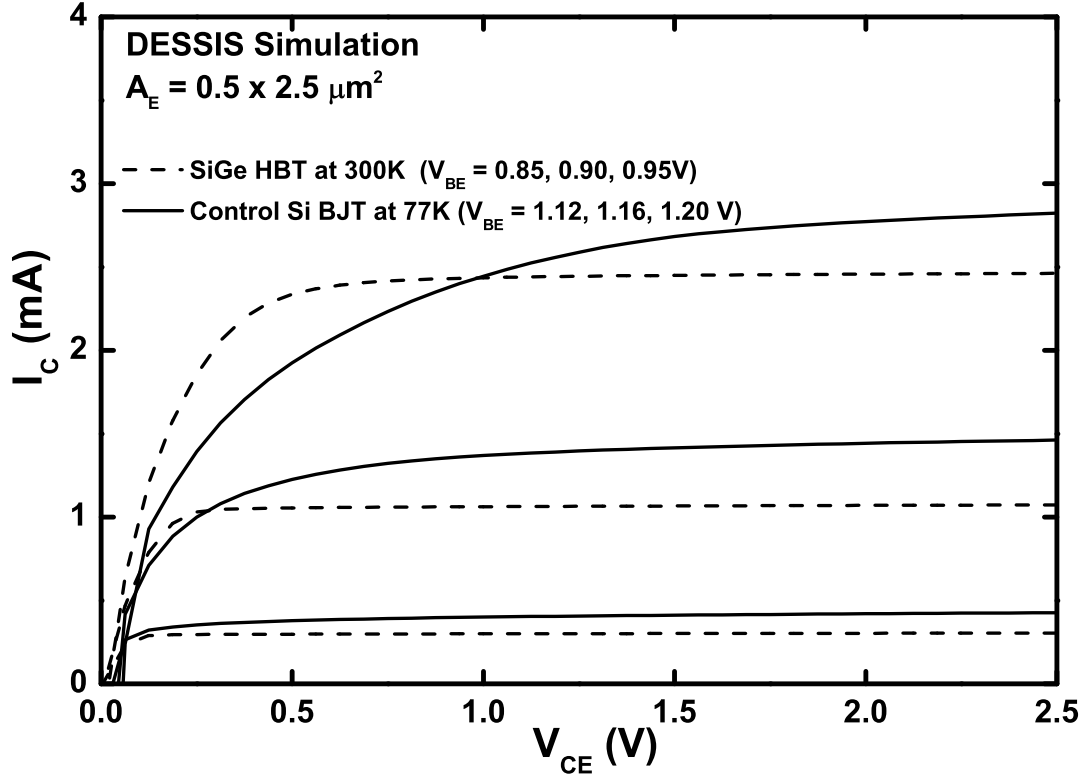
Both the NDR and the  $I_C$  kink effect can be qualitatively captured using 2-D DESSIS simulations. The 2-D doping and Ge profiles and device mesh were based on SIMS data and the actual transistor layout dimensions. The simulation parameters have been carefully calibrated to fit the measured DC and AC characteristics of the transistor across temperature. 77 K was the lowest simulation temperature used because robust convergence at high-injection could not be easily attained below this temperature.



**Figure 12:** DESSIS simulations of the SiGe HBT at 77 K, including (a) forced- $V_{BE}$  output characteristics showing NDR, and (b) Gummel characteristics and the resultant  $g_m$  where the  $I_C$  kink effect occurs.

Fig. 12 shows the simulation results of the SiGe HBT at 77 K, clearly showing the NDR, the  $I_C$  kink effect, and the  $g_m$  spike. No such effects are found in the simulations

of the control Si BJT at 77 K or the SiGe HBT operating at 300 K (Fig. 13). Turning off the hydrodynamic model in DESSIS makes little difference in the result, which indicates that the NDR effects can be explained within the traditional drift-diffusion framework (turning off carrier freezeout in the simulation produces the same trends, showing that carrier freezeout does not play a strong role either).

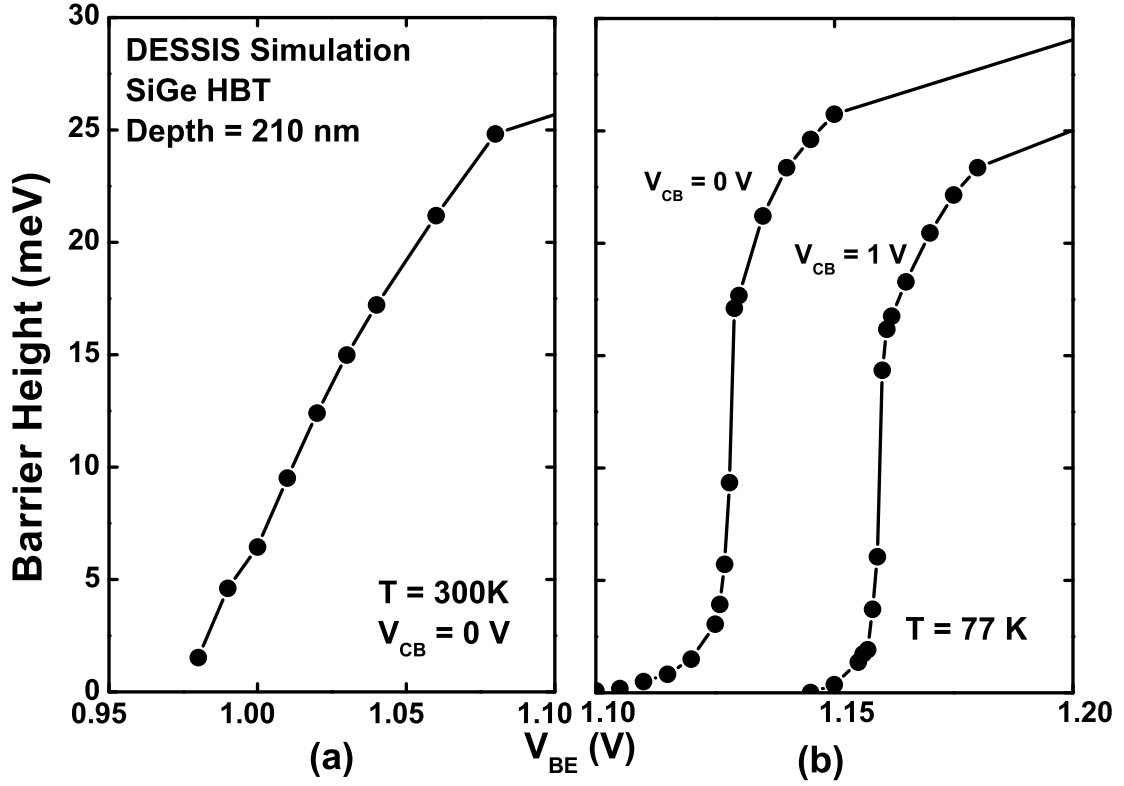


**Figure 13:** DESSIS simulations of the forced- $V_{BE}$  output characteristics of the SiGe HBT at 300 K, and the control Si BJT at 77 K.

#### 2.4.1 Collector-Current Kink Effect

The barrier height and  $n$  in the base at the onset of HBE as functions of  $V_{BE}$  are shown in Fig. 14 and 15. The snapshot of these two parameters are taken at the position of the cut line indicated in Fig. 10.

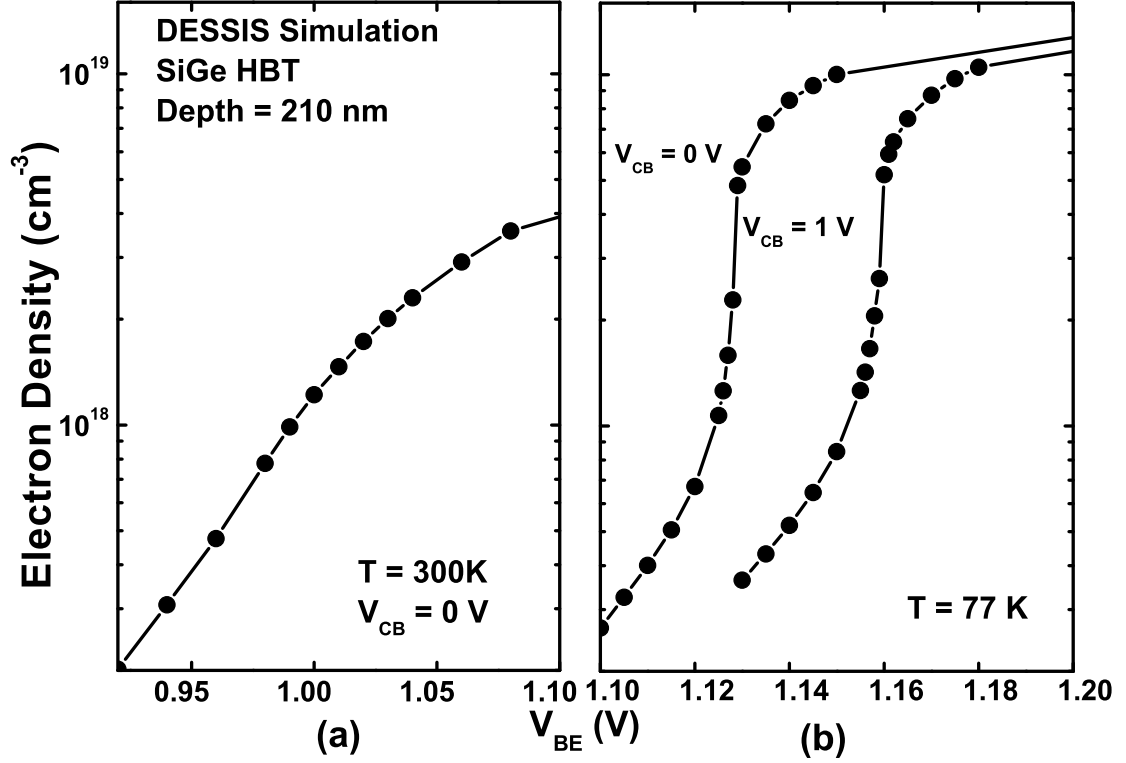
The SiGe HBT operating at 300 K shows a gradual increase in the barrier height



**Figure 14:** Simulated heterojunction barrier height as functions of  $V_{CE}$  for the SiGe HBT at (a) 300 K ( $V_{CB} = 0.0\text{ V}$ ) and (b) 77 K ( $V_{CB} = 0.0$  and  $1.0\text{ V}$ )

and the resultant  $n$  with increasing  $V_{BE}$ ; however, both the barrier height and  $n$  increase far more rapidly at the onset of HBE in the device operating at 77 K. This is expected in the context of our proposed mechanism, where the positive feedback of the barrier effect is demonstrated to be greatly enhanced at very low temperatures. Changing the CB bias from 0 V to 1 V helps push the sudden onset of the barrier effect to a higher  $V_{CB}$  (from 1.128 V to 1.159 V in this case), as shown in Fig. 14. However, both CB biases at 77 K result in an extremely rapid turn-on of the barrier effect, producing a sudden increase in the barrier height and the drastic increase in charge accumulation, which is clearly different than at room temperature.

The resultant sharp increase in  $n$  causes an extra drift current component that adds up to the total collector current, consequently producing the  $I_C$  kink and the



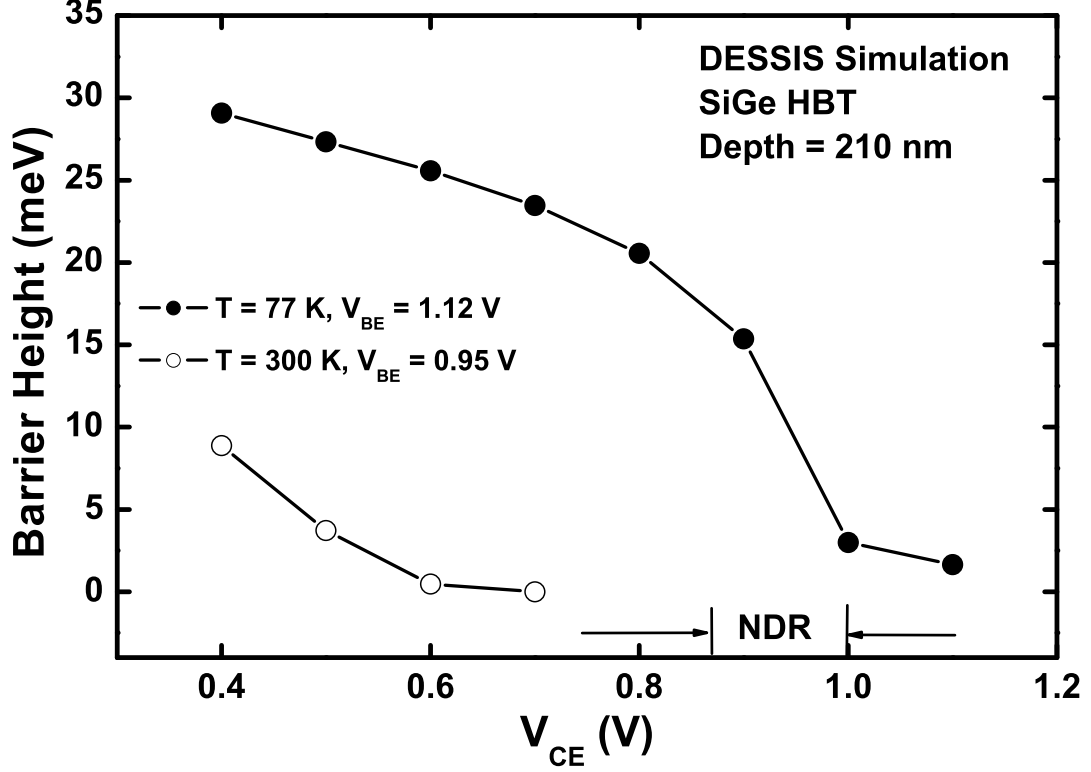
**Figure 15:** Simulated electron density as a function of  $V_{BE}$  for the SiGe HBT at (a) 300 K ( $V_{CB} = 0.0$  V) and (b) 77 K ( $V_{CB} = 0.0$  and 1.0 V). The snapshot is taken at the position of the cut line indicated in Fig. 10.

$g_m$  spike present in the Gummel characteristics of the device. When  $V_{BE}$  further increases to the values higher than the kink effect, it takes much more increase in the total current density to induce the same amount of barrier height growth, as the barrier height is eventually limited by bandgap difference between Si and SiGe. Therefore the condition of positive feedback is valid only at the onset of HBE, after which  $I_C$  clamps (saturates), as is generally true for HBE.

#### 2.4.2 The NDR Picture

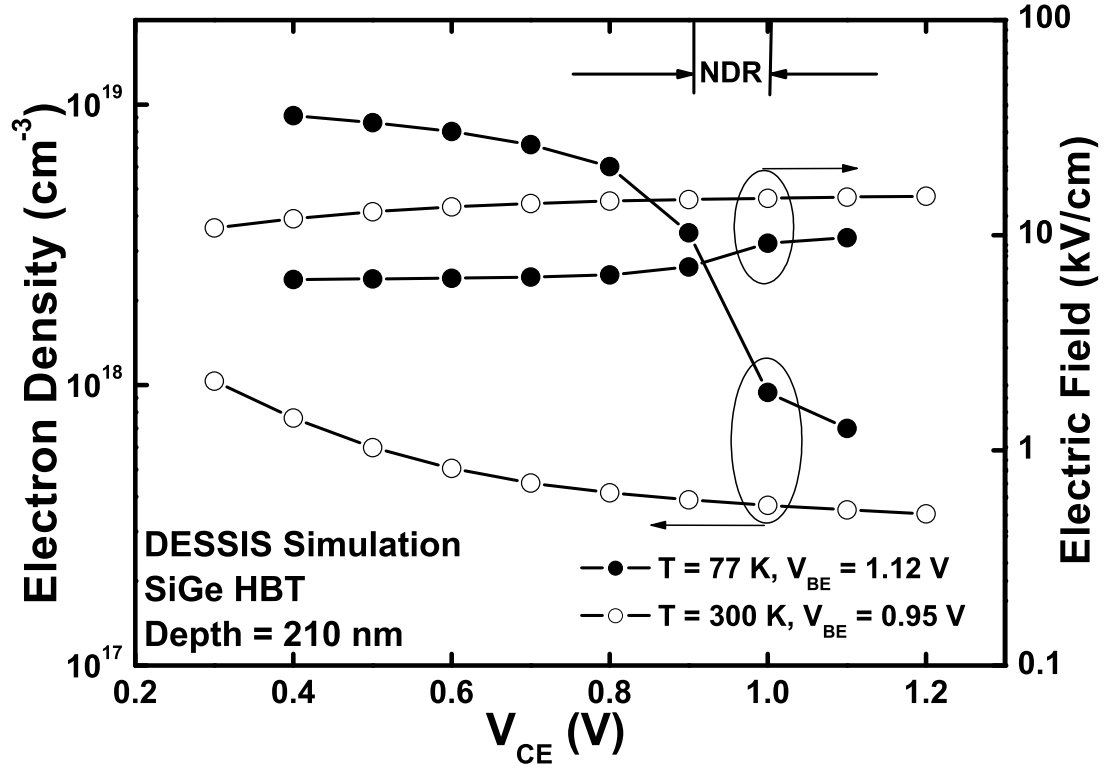
To shed further light on the mechanism underlying the NDR seen in the forced- $V_{BE}$  output characteristics, two more physical parameters need to be examined in the analysis. Fig. 16-18 provide a comprehensive picture of  $n$ , the heterojunction barrier

height, the magnitude of the electric field ( $E$ ), the electron drift current component, and the total electron current component, extracted from a family of points on the forced- $V_{BE}$  characteristics (at both 77 K and 300 K) at the position of the cut line indicated in Fig. 10.



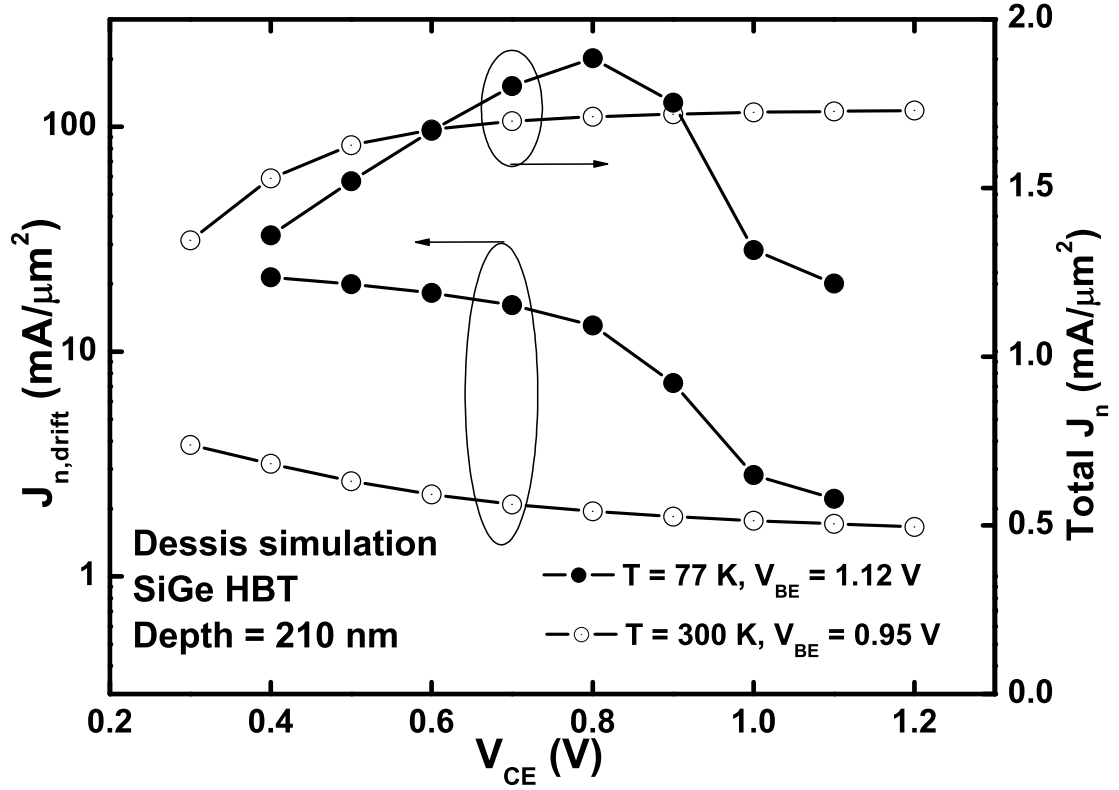
**Figure 16:** Simulated heterojunction barrier height as a function of  $V_{CE}$  for the SiGe HBT at 77 K and 300 K. The snapshot is taken at the position of the cut line indicated in Fig. 10.

Increasing  $V_{CE}$  decreases the base electron accumulation in two ways: a higher  $V_{CB}$  changes the Shockley boundary condition and the resultant electron concentration, and in addition, higher  $V_{CB}$  decreases the heterojunction barrier height (Fig.16) with a resultant change in carrier concentration (Fig. 17). Consequently, the electron drift current component decreases with increasing  $V_{CE}$ , as shown in Fig. 18, even though the electric field increases steadily (Fig. 17). At the same time higher  $V_{CE}$  changes the electron gradient and the resultant electron diffusion current density increases.



**Figure 17:** Simulated electron density and electric field as functions of  $V_{CE}$  for the SiGe HBT at 77 K and 300 K. The snapshot is taken at the position of the cut line indicated in Fig. 10.

These two contributions both affect the total electron current density. At room temperature, when  $n$  is smaller and only slightly changes with increasing  $V_{CE}$ , the change of the drift current component is expected to be smooth, as shown in the simulated 300 K curve in Fig. 18. Considering the gradually increasing diffusion current component, the total electron current density is an increasing function of  $V_{CE}$  (Fig. 18). However, at cryogenic temperatures (e.g., 77 K) there is a bias region over which a rapid barrier height decrease occurs because of the enhanced positive feedback in HBE (Fig. 16), producing a significant decrease in  $n$  (Fig. 17), and a sharp turn-off of the HBE at that bias region. The barrier effect becomes much more sensitive to  $V_{CB}$  in this region. The relative decrease in  $n$  is more rapid than the increase in the electric field (Fig. 17), and the plot of the resultant drift current density thus shows



**Figure 18:** Simulated drift component of the electron current density and the total electron current density as functions of  $V_{CE}$  for the SiGe HBT at 77 K and 300 K. The snapshot is taken at the position of the cut line indicated in Fig. 10.

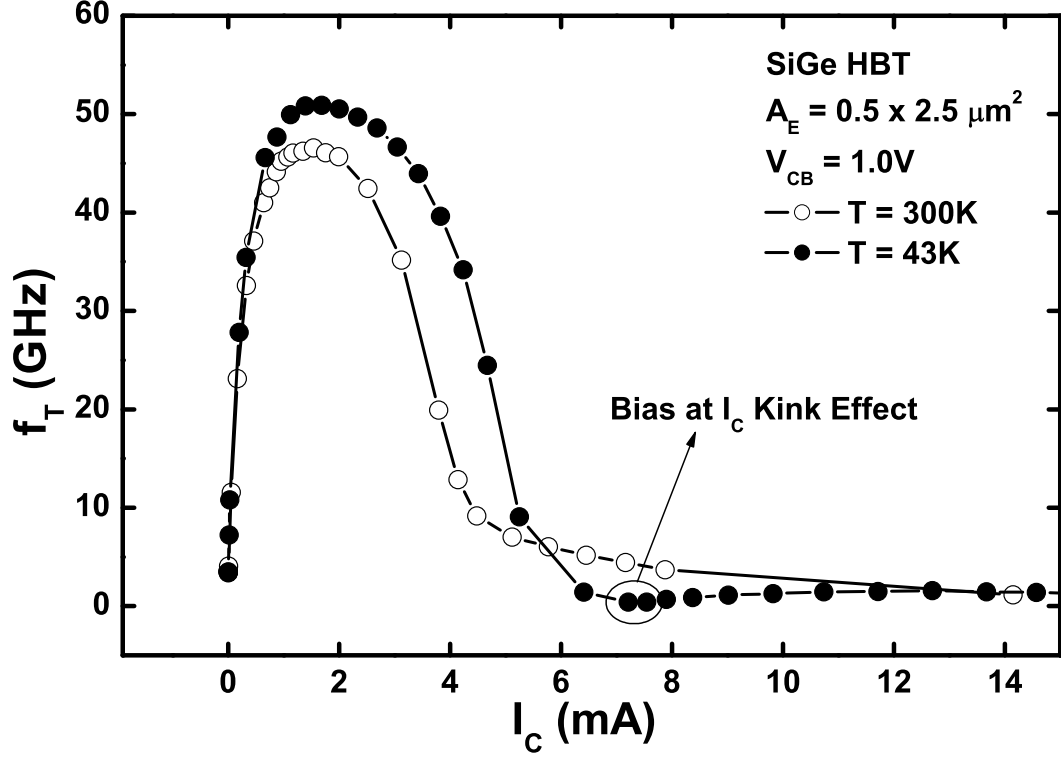
an unusual sudden drop at 77 K, directly resulting in the fact that the total electron current reaches a peak before decreasing (Fig. 18). This produces the region of NDR in the total  $I_C$ , consistent with the measured data.

## 2.5 AC Consequences

These novel effects also influence the AC performance of the device at cryogenic temperatures. In contrast to the situation at 300 K, where  $f_T$  decreases monotonically with increasing  $I_C$  above peak  $f_T$ , the extra accumulation of charges in the base due to HBE at cryogenic temperatures causes an unusual “dip” in the  $f_T$ - $I_C$  characteristics (Fig. 19). That is, the AC degradation is larger than one would otherwise expect, and partially offsets the generally favorable impact of cooling on SiGe HBT dynamic



performance. Fortunately, however, for most circuit applications, this dip occurs well above operational bias current densities.



**Figure 19:** Measured cut-off frequency ( $f_T$ ) vs. collector current ( $I_C$ ) for the SiGe HBT at 77 K and 300 K. The 77 K plot shows an  $f_T$  “dip” in accordance with the  $I_C$  kink effect in its DC characteristics.

## 2.6 Collector Doping and Technology Scaling

To quantitatively investigate the influence of collector doping and technology scaling, a normalized transconductance per unit current parameter ( $g'_m$ ), which can be easily extracted from the Gummel characteristics of the device, and is defined as,

$$g'_m = \frac{g_{m,peak}}{I_{C,barrier}} \quad (15)$$

where  $g_{m,peak}$  is the peak  $g_m$  value in the  $g_m$  spike (using a  $V_{BE}$  step of 2 mV) and  $I_{C,barrier}$  stands for the threshold collector current at which HBE turns on. Substituting (13) and (14) in Chapter III into (15), one obtains,

$$g'_m = \frac{1}{I_{C,barrier}} \cdot \frac{\partial I_C}{\partial V_{BE}} = \frac{1}{I_{C,barrier}} \cdot \left( \frac{\partial I_n}{\partial V_{BE}} + \frac{\partial I_p}{\partial V_{BE}} \right) \quad (16)$$

Consequently,

$$g'_m = \frac{1}{J_{C,barrier}} \cdot \frac{\partial \left[ q \cdot \mu_n \cdot n \cdot E + q \cdot \mu_n \cdot \nabla \left( \frac{nkT}{q} \right) + J_p \right]}{\partial V_{BE}} \quad (17)$$

Or,

$$g'_m \approx \frac{1}{J_{C,barrier}} \cdot \frac{\partial (q \cdot \mu_n \cdot n \cdot E)}{\partial V_{BE}} = \frac{1}{n} \cdot \frac{\partial n}{\partial V_{BE}} + \frac{1}{E} \cdot \frac{\partial E}{\partial V_{BE}} \quad (18)$$

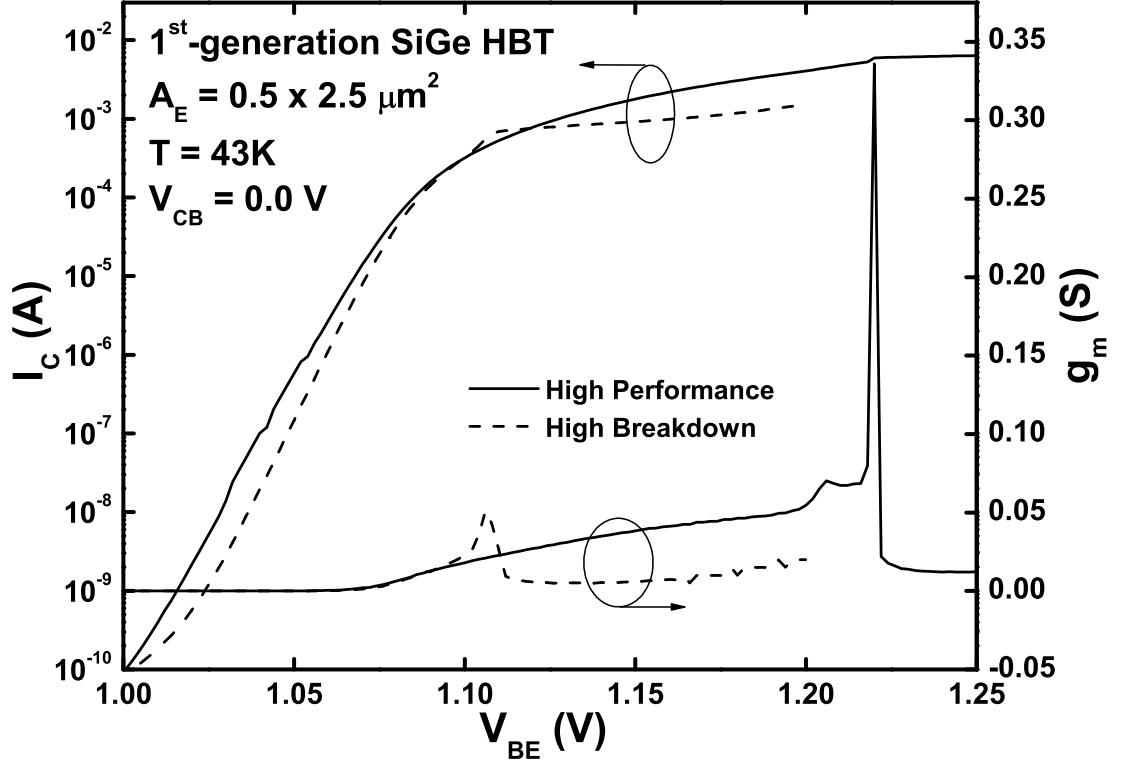
At the onset of HBE at deep cryogenic temperatures,  $\frac{1}{E} \cdot \frac{\partial E}{\partial V_{BE}} \ll \frac{1}{n} \cdot \frac{\partial n}{\partial V_{BE}}$  because of the enhanced positive feedback that rapidly accumulates electrons in the base. As a result, (18) can then be simplified as,

$$g'_m \approx \frac{1}{n} \cdot \frac{\partial n}{\partial V_{BE}} \Big|_{V_{BE}=V_{BE,barrier}} \quad (19)$$

where  $V_{BE,barrier}$  is the  $V_{BE}$  at which HBE turns on. Therefore,  $g'_m$  represents a useful parameter that quantifies the greatly enhanced charge accumulation due to barrier effect, and enables a comparison among different devices or in principle even different technology nodes, independent of parameters such as electron velocity, doping profile, and the depth of SIC region, all of which are strongly profile dependent and typically unknown.

The impact of the collector doping on these HBE-induced cryogenic effects can be experimentally investigated by comparing the high-performance (HP) SiGe HBT with its high-breakdown (HB) SiGe HBT counterpart, both of which exist side-by-side on the same wafer, because the only difference between these two devices lies in the lower collector doping level of the HB device (i.e., the Ge profile is the same).

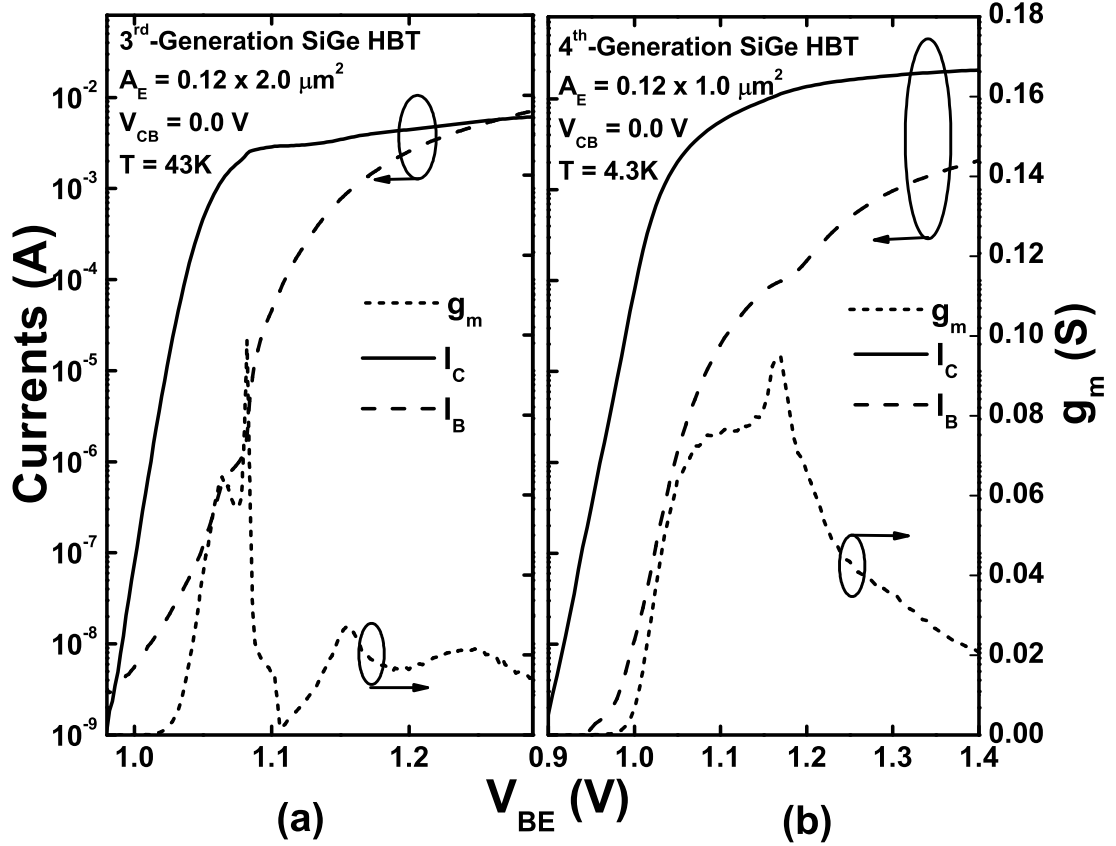
The  $g_m$  and the  $I_C$  characteristics of the first-generation HP and HB SiGe HBTs are shown in Fig. 20, and  $g'_m$  are shown in the upper two curves of Fig. 22. The lower collector doping of a HB device yields an earlier onset of both the classical Kirk effect and HBE, and thus is naively expected to have a larger impact on the device Gummel characteristics at low temperatures. Therefore, the HB device with a lower SIC doping is expected to show a larger  $g'_m$  than the HP device, in agreement with the data.



**Figure 20:** Measured Gummel characteristics and transistor transconductance ( $g_m$ ) at  $V_{CB} = 0$  V for the HB and HP first-generation SiGe HBTs at 43 K.

With regard to technology scaling, the  $I_C$  kink effect and the  $g_m$  spike also exist in both third-generation ( $f_T = 200$  GHz at 300 K) and state-of-the-art fourth-generation prototype ( $f_T = 350$  GHz at 300 K) SiGe HBTs operating at deep cryogenic temperatures (Fig. 21), but the impact is suppressed in magnitude compared to that in

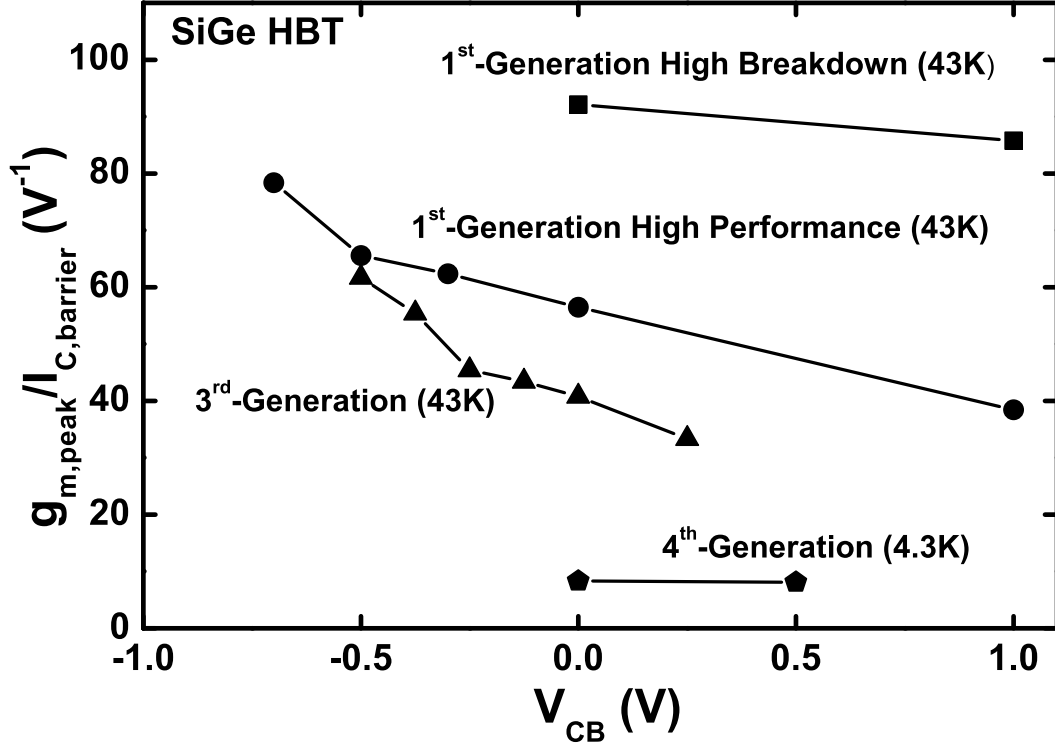
first-generation SiGe HBTs. In fact, the  $g_m$  spike in the fourth-generation SiGe HBT is only observable at extremely low temperatures ( $< 10$  K). Additionally, the NDR in the forced- $V_{BE}$  output characteristics is entirely buried by the increasing electron velocity with the increase of  $V_{CE}$ , and cannot be clearly observed, even at 5.4 K. The suppression of these effects with technology scaling is expected, given that the collector regions of the scaled transistors are naturally much more heavily doped, partially mitigating HBE and its impact on device performance.



**Figure 21:** Measured Gummel characteristics and transistor transconductance ( $g_m$ ) at  $V_{CB} = 0$  V for (a) the third-generation SiGe HBT (200 GHz at 300 K) at 43 K, and (b) the fourth-generation SiGe HBT (350 GHz at 300 K) at 4.3 K.

Fig. 22 summarizes the  $g'_m$  of the first-, third-, and fourth-generation HP SiGe HBTs and the first-generation HB SiGe HBTs, highlighting this trend of suppression with the advancement in device speed and technology node. In addition, one can

see from the figure that higher  $V_{CB}$  causes lower  $g'_m$  as HBE is suppressed at higher CB bias. One can thus conclude that the charge accumulation due to the positive feedback in HBE at cryogenic temperatures can be suppressed as long as the HBE is mitigated by higher SIC doping, higher CB bias, and careful technology scaling. This is clearly good news.

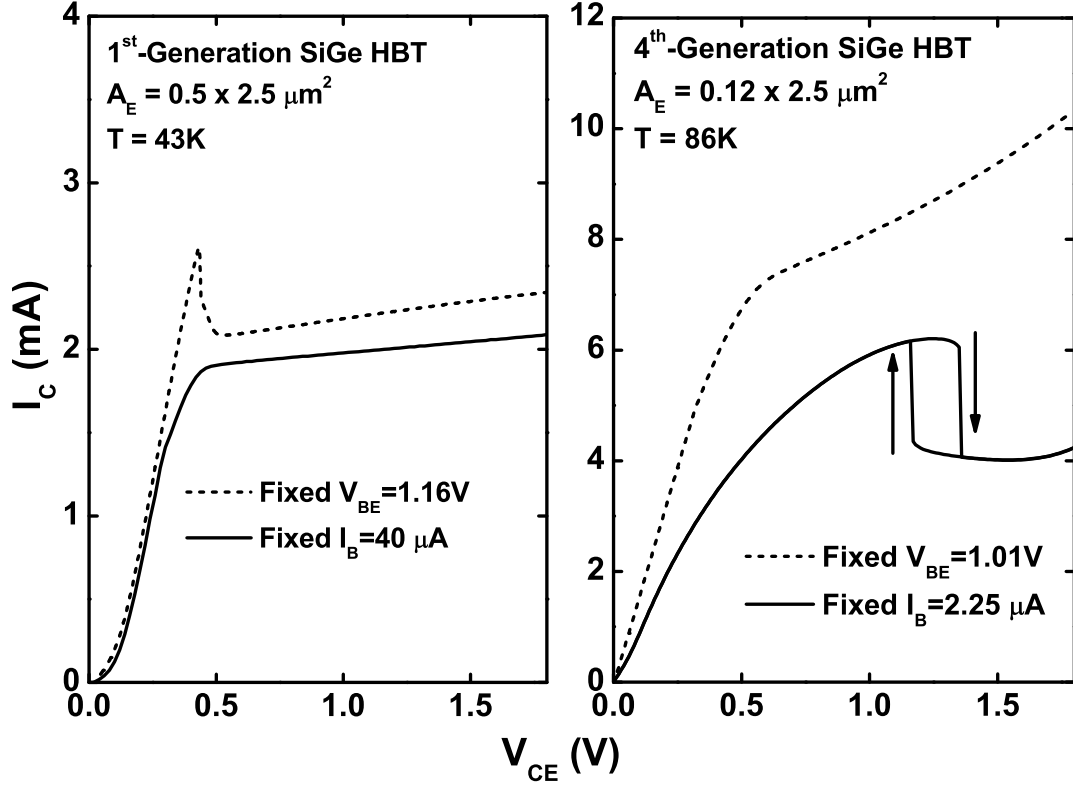


**Figure 22:** Measured peak unit transconductance ( $g'_m$ ) as a function of  $V_{CB}$  for the first- (both HP and HB), third-, and fourth-generation SiGe HBTs.

## 2.7 Circuit Implications

The NDR presented in the previous sections is observed in the forced- $V_{BE}$  output characteristics of the first-generation SiGe HBTs, while there is no NDR in the forced- $I_B$  output characteristics. Interestingly, in [44] an NDR effect and an unusual “hysteresis” behavior in the forced- $I_B$  output characteristics of the fourth-generation SiGe HBTs at cryogenic temperatures was reported. However, no NDR can be seen

in the forced- $V_{BE}$  output curves of those devices. The interesting contrast between these two related phenomena is summarized in Fig. 23.



**Figure 23:** Comparison of the different NDR effects in the first-generation and the state-of-the-art fourth generation SiGe HBTs. NDR is observed in the forced- $V_{BE}$  output characteristics of the first-generation SiGe HBTs, but in the forced- $I_B$  output characteristics of the fourth-generation devices together with a hysteresis behavior.

The NDR in the forced- $I_B$  output curves results from the influence of tunneling through EB junction  $\Delta I_E$ . The  $\Delta I_E$  can be divided into the non-ideal collector leakage current ( $\Delta I_C$ ), as described in [9], and the associated base recombination current ( $\Delta I_B$ ) that recombine with the holes in the valence band [44]. The weakly temperature dependent  $\Delta I_B$  comprises a large portion in the total base current at extremely low temperatures when the conventional base current component is small, and this  $\Delta I_B$  can then be expressed as,

$$\Delta I_B \propto \overline{U_{SRH,p}} \approx \frac{(r_p \bar{p} + s_p) f(V_{BE})}{r_n \bar{n} + r_p \bar{p} + s_n + s_p} \quad (20)$$

where  $\overline{U_{SRH,p}}$  is the effective hole recombination rate,  $\bar{n}$  ( $\bar{p}$ ) is the effective electron (hole) density in the base,  $s_n$  ( $s_p$ ) is the trap electron (hole) generation rate,  $r_n$  ( $r_p$ ) is the electron (hole) recombination rate, and  $f(V_{BE})$  is the tunneling-induced electron transition that is an increasing function of  $V_{BE}$  [44].

During the measurement of the transistor output characteristics, the heterojunction barrier decreases when  $V_{CE}$  increases, as indicated in the above discussion, and as a result,  $n$  (and thus  $\bar{n}$ ) decreases. This will greatly increase  $\Delta I_B$ , according to Eq. (20). In the forced- $I_B$  measurement, as  $V_{CE}$  is gradually increased,  $V_{BE}$  needs to shift to a smaller value to compensate for this increase in the total  $I_B$ , and consequently there is an induced NDR in the final output characteristics. The associated  $I_B$  “dip” in the Gummel characteristics can be explained using the same theory. Interested readers should refer to the complete discussion in [44] for additional details. The observed “hysteresis” behavior is the result of a bifurcation of the  $I_B - V_{BE}$  solutions in this bias region – that is, there are two  $V_{BE}$ ’s that can sustain the same  $I_B$ : the higher  $V_{BE}$  with a larger traditional  $I_B$  component, and the lower  $V_{BE}$  with a larger  $\Delta I_B$ .

Both NDR effects are cryogenically-enhanced phenomena, and are closely related to the electron accumulation in the base due to HBE. Therefore they are both observed in the bias region where HBE dominates the device response. In fact, one can see in the right half figure of Fig. 21 that both the  $g_m$  spike and the  $I_B$  “dip” occur under the same bias conditions. However, they are fundamentally different mechanisms.

From a circuit application perspective, one NDR effect is operative only in the constant-voltage input bias mode, while the other NDR effect is operative only in the constant-current input bias mode. For the constant-voltage input mode NDR, the electron accumulation in the base forms an additional drift current component that adds to the total collector current, while in the constant-current input mode NDR, the HBE-induced electron accumulation in the base influences the base current

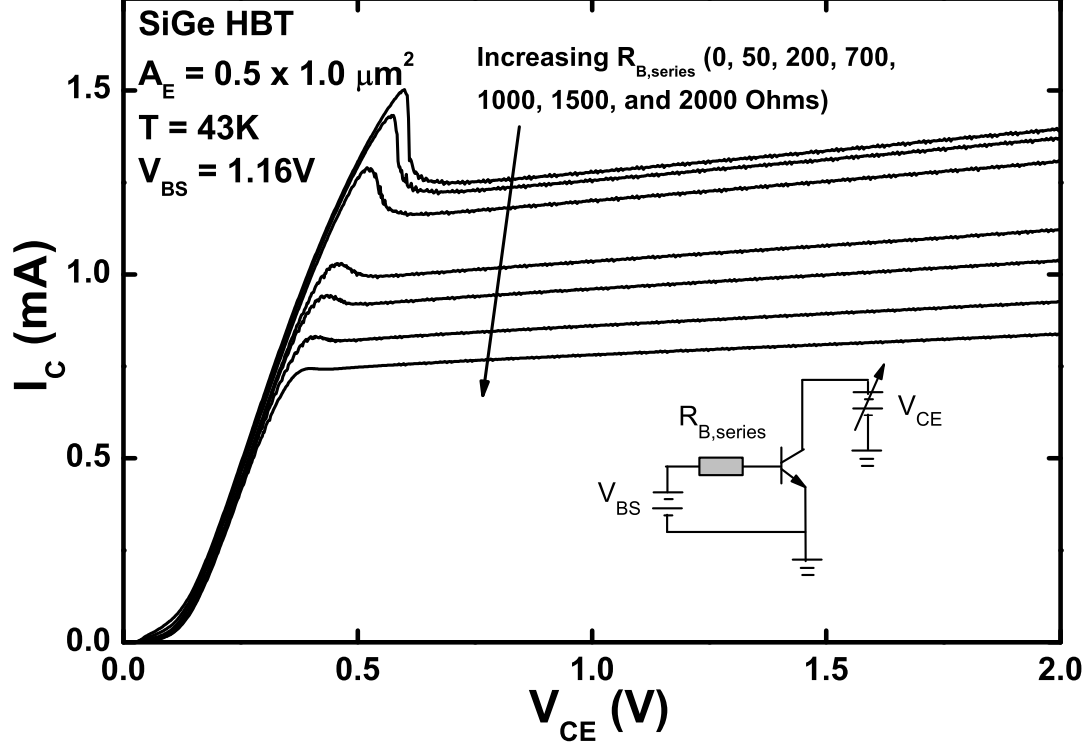
only by decreasing  $\Delta I_B$ , as shown in Eq. (20). Moreover, the offsetting influence of HBE on  $\Delta I_B$  and the classical  $I_B$  component renders a hysteresis in the constant-current input mode of operation, while but no hysteresis or bifurcation expected in the constant-voltage input mode. Finally, the constant-voltage input NDR is proven to be mitigated with natural technology scaling, as stated before, since the more advanced technology nodes naturally have more aggressive collector doping, which acts to smooth the sudden increase of base charge induced by HBE. On the other hand, the constant-current input NDR becomes increasingly significant as technology advances (as evidenced by the slight NDR in the third-generation SiGe HBTs as shown in [7] and the obvious NDR in the fourth-generation), because the thinner EB space charge region resulted from the higher doped emitter and base greatly enhances the tunneling process which drives the phenomena.

Traditionally, circuit designers rarely worry about the differences between the constant-current input and constant-voltage input modes of operation for bipolar transistors. Either set of output characteristics, regardless of forced- $I_B$  or forced- $V_{BE}$  bias, is typically regarded as sufficient for understanding circuit design [136], [137]. While a difference in Early voltage with respect to the two different modes of operation was reported in [138] for Si BJTs, more serious circuit-relevant concerns were highlighted by *Joseph et al.*, who reported the influence of neutral base recombination on the temperature dependence of Early voltage in SiGe HBTs [139].

There are in fact circuit-relevant examples which invoke a forced- $I_B$  or a forced- $V_{BE}$  mode of operation of the transistors, including high-source resistance current sources and low-source resistance current sources [139]. In reality, most circuits in real applications operate in a state lying somewhere between these two extremes – that is, they are not driven by an ideal current source or an ideal voltage source, but something in between. Equivalently, a series resistance  $R_{B,series}$  can be connected between the base and the voltage source. The constant-voltage input mode is then



represented by  $R_{B,series} = 0$ , while the constant-current input mode is represented by  $R_{B,series} = \infty$ , the two extremes of most practical circuits.

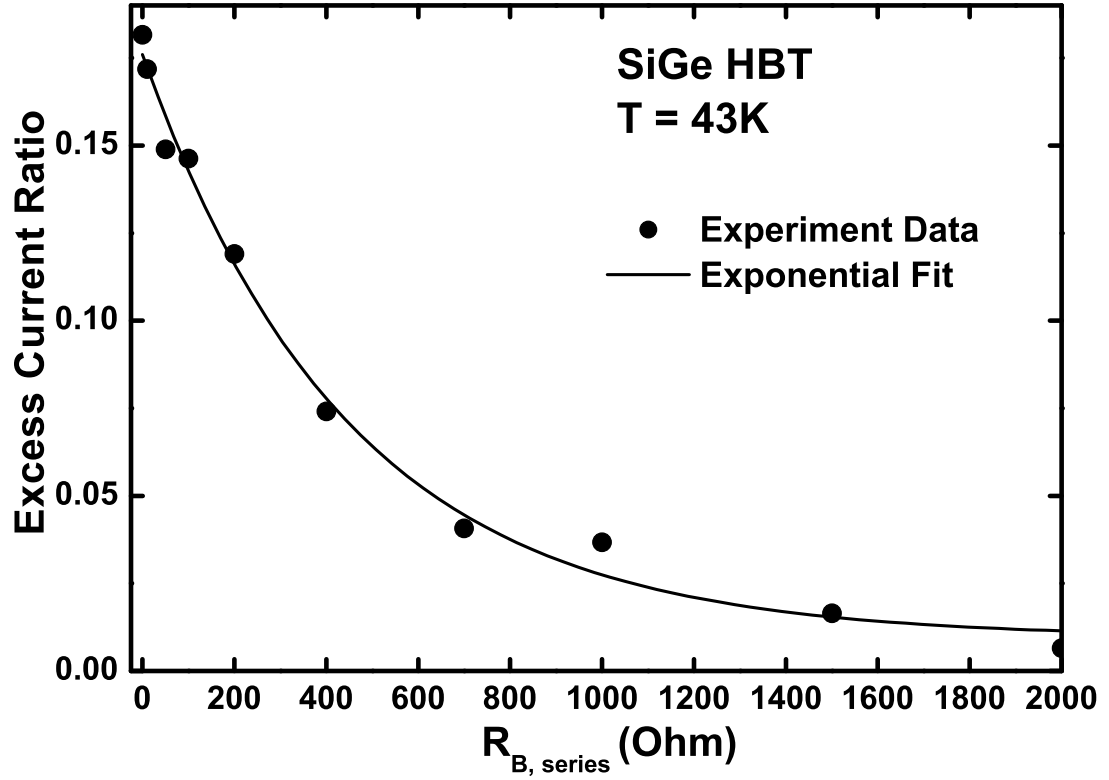


**Figure 24:** The measured forced- $V_{BE}$  output characteristics of SiGe HBT at 43 K and at  $V_{BS} = 1.16$  V by changing the value of the base series resistance.

For better understanding of how the present NDR potentially could affect real circuit designs, a complete set of output characteristics were measured using the circuit configuration shown in the inset of Fig. 24. The output characteristics gradually approach the pure forced- $I_B$  output characteristics, and the NDR weakens as  $R_{B,series}$  increases from 0 to 2 k $\Omega$  (Fig. 24). One can also observe that the NDR almost disappears when  $R_{B,series} = 2$  k $\Omega$ . Quantitatively, the associated  $ECR$  decays exponentially with  $R_{B,series}$ , as shown in Fig. 25, which can be modeled as,

$$ECR(R_{B,series}) = ECR_0 + B \cdot e^{(-R_{B,series}/R_0)} \quad (21)$$

where the fitting parameters  $ECR_0 = 0.0095$ ,  $B = 0.1665$ , and  $R_0 = 449.1\Omega$ .



**Figure 25:** The measured excess current ratio at 43 K as a function of the base series resistance.

The output impedance of the previous stage of the circuit or the internal resistance of a voltage source can be regarded as  $R_{B,series}$  in real circuits and can obviously have a strong influence on the magnitude of the HBE-induced NDR effect. Circuit designers worried about such effects should incorporate the output impedance of their input circuit into their designs to carefully assess the impact of these novel NDR effects. Resultant NDR-induced circuit instabilities, in this context, may indeed be of concern to circuits biased in this region, and potential circuits of concern would include the output stages of analog circuits, current sources, and high-current drivers. Careful compact modeling is obviously warranted, and we note that existing compact models for SiGe HBTs (e.g., VBIC, MEXTRAM, or HICUM) do not at present account for such unique NDR effects.

## 2.8 *Summary*

A detailed investigation of a new NDR effect and a novel collector-current kink effect in cryogenically-operated SiGe HBTs has been presented in this chapter. It has been shown that the additional drift current due to the charge accumulation induced by the enhanced positive feedback in HBE causes these effects. 2-D DESSIS simulations are used to confirm the proposed mechanism. The enhanced positive feedback at extremely low temperatures influences the AC performance of the device by producing an  $f_T$  “dip” in the  $f_T - I_C$  characteristics. Technology scaling is shown to partially mitigate these effects. Circuit designers working in these environmental regimes should pay close attention to potential NDR-induced instabilities.

## CHAPTER III

# BASE PROFILE OPTIMIZATION FOR RF LUNAR APPLICATIONS

As more SiGe electronic sub-systems using first-generation SiGe HBTs for sensing, actuation, and robotics on the Moon are under development. Meanwhile, a logical additional application of SiGe would be in cryogenically-enabled wireless S-, Ka-, or K-band communication systems. We have previously reported that the AC performance of the “control” (i.e., currently in manufacturing) first-generation SiGe HBTs degrades below 100 K, potentially limiting the application of these transistors in the RF domain. Therefore, optimizing the high-frequency performance of SiGe HBTs for operation to 43 K has become particularly important, and is addressed here for the first time.

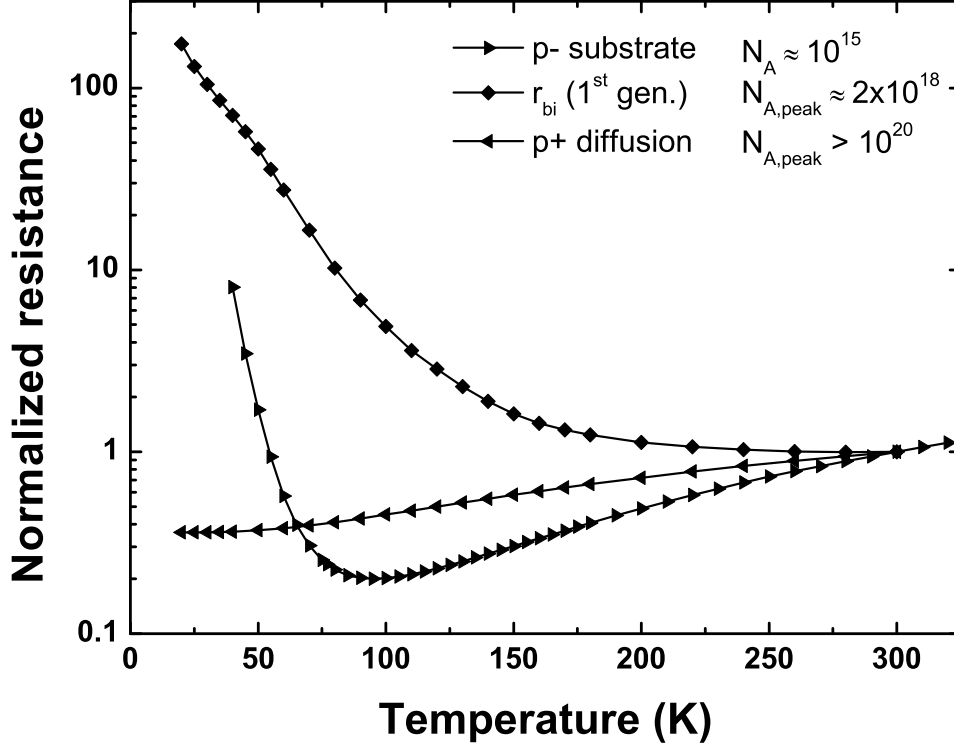
Based on the observed performance of the control first-generation SiGe HBTs, intentional changes to the germanium and the base doping profiles were made and evaluated from 43 to 300 K. Enhanced  $f_T$  and  $f_{max}$  are observed in the two optimized devices at cryogenic temperatures. Further analysis indicates that these new profiles suppress heterojunction barrier effects and carrier freezeout at cryogenic temperatures, resulting in improved performance at 43 K. Finally the breakdown voltages are measured in all three devices over temperature to ensure their applicability for RF circuit blocks.

### ***3.1 Base Carrier Freezeout***

A closer look at the degradation of the device performance of the control SiGe HBTs at 43 K reveals that the heterojunction barrier effect (HBE) and carrier freezeout in

the base are limiting device performance at deep cryogenic temperatures. HBE is triggered via Kirk effect and was investigated extensively in the previous chapter. When the mobile charge density in the CB depletion region exceeds the local ionized doping density, the resultant electric field in this region starts to collapse and the neutral base region pushes out into the selectively implanted collector. This pushed-out base exposes the SiGe-Si heterojunction inducing a conduction band barrier. HBE then degrades the  $f_T$ ,  $g_m$ , and  $\beta$  at high injection, and is greatly enhanced at cryogenic temperatures because they are band edge phenomena and hence thermally activated.

The other big concern at cryogenic temperatures is carrier freezeout (or incomplete ionization) [140], [141]. The carrier transport properties of Si (or SiGe) depend on the doping level, and the issue of carrier freezeout is eliminated if the semiconductor is doped above Mott-transition [142]. To understand why higher base doping is beneficial, one needs to explore the difference between intrinsic and extrinsic silicon: significantly smaller thermal excitation is needed to generate carriers in doped silicon [143], [144]. Though a weak function of temperature itself, the dopant ionization energy in lightly-doped silicon is harder to be reached at lower temperatures, and thus carrier freezeout becomes significant, causing lower conductivity during cryogenic operation. However, when the dopants are so close that their wave-functions overlap, impurity bands will form and no ionization energy is required to conduct current (the so-called Mott-transition). The critical doping level for Mott-transition in Boron-doped silicon is about  $4 \times 10^{18} \text{ cm}^{-3}$  [145]. Therefore, carrier freezeout in Si (or SiGe) doped below this level is detrimental to the cryogenic performance of SiGe HBTs, especially to the base resistance ( $R_B$ ), which can obviously impact  $f_{max}$ . Other freezeout-related effects, such as minority carrier trapping, can also degrade  $f_T$  at low temperatures [146].



**Figure 26:** Temperature dependence of the normalized base sheet resistance of a first-generation SiGe HBT in comparison with the normalized resistance of other p-type Si/SiGe layers (data provided by K. Moen).

The temperature dependence of carrier mobility and dopant ionization level mutually affect the sheet resistance of extrinsic silicon. Fig. 26 shows the measured resistance of p-type silicon *vs.* temperature for a range of doping concentrations, normalized to 300 K. The data were provided by Kurt Moen in [147]. For doping concentrations below the Mott transition (e.g., in the p-substrate and base region of a first-generation NPN SiGe HBT), the sheet resistance increases at lower temperature. The intrinsic base sheet resistance ( $R_{bi}$ ) of the first-generation SiGe HBT is more than ten times of its 300-K value when temperature  $< 50$  K. For highly-doped  $p^+$  diffusion

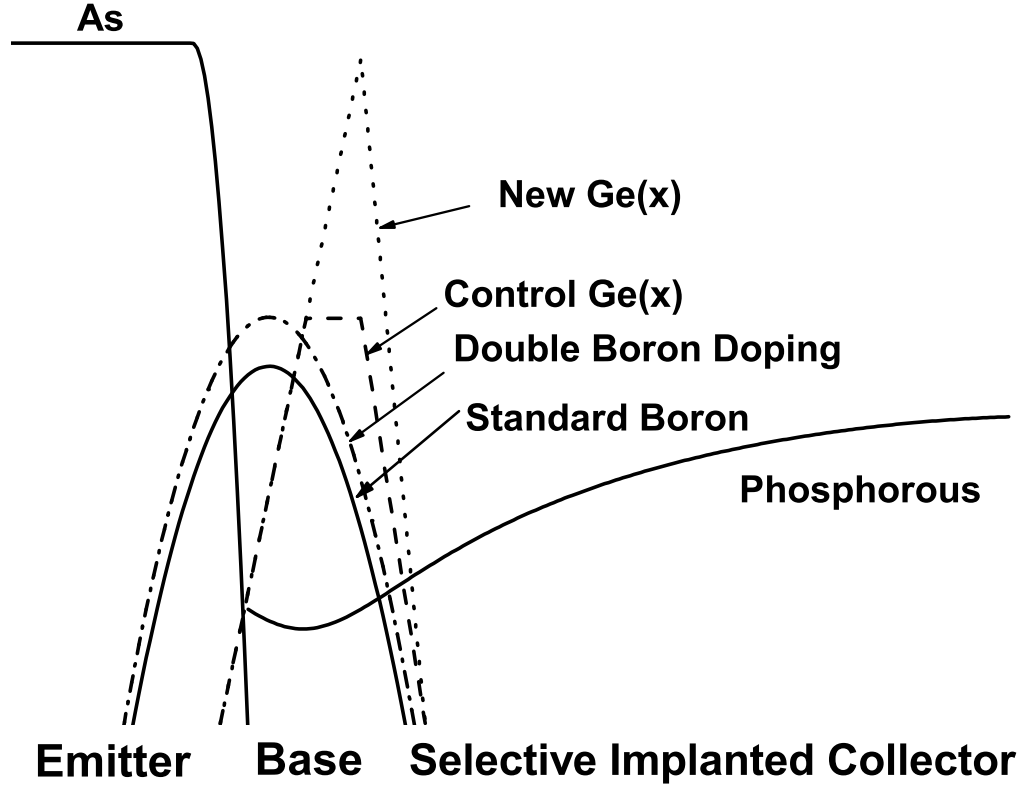
layer dope above Mott-transition, incomplete ionization is negligible across temperature, and the change in base resistance across temperature is mainly driven by the carrier mobility.

### 3.2 *Base Profile Design*

The solid lines in Fig. 27 illustrate the doping and Ge profiles of a control first-generation SiGe HBT featuring a trapezoidal Ge profile (the dashed line) and a metallurgical base width of 80 nm. The peak boron doping in the epitaxial base is about  $2 \times 10^{18} \text{ cm}^{-3}$ . A recent study has shown that maintaining the front-side Ge gradient instead of extending the Ge into the depletion region of the CB junction is more efficient in suppressing HBE at cryogenic temperatures [148]. Therefore, a triangle Ge profile was designed here with the same Ge slope but higher peak Ge fraction, and labeled “*new Ge(x)*” profile (the dotted line in Fig. 27).

As indicated in the previous section, increasing the base doping can gradually degenerate silicon from semiconducting behavior to quasi-metallic behavior over temperature. Here we have doubled the base doping so that the peak boron concentration is slightly above the Mott-transition. The actual neutral base region of the higher doped base is now wider and the collector doping is effectively higher (Fig. 27). A minor degradation in  $f_T$  and  $\beta$  at 300 K can be expected because of the higher base doping and the associated higher  $\tau_B$  (via lower mobility). In addition, the intrinsic base resistance ( $R_{bi}$ ) can be greatly reduced with higher doping; it does not, however, dominate the total  $R_B$  at 300 K, as the extrinsic base resistance  $R_{bx}$  is of similar magnitude as  $R_{bi}$ . However,  $R_{bi}$  dominates the total  $R_B$  at low temperature when freezeout occurs in the intrinsic base, but not in the extrinsic base, where the heavy doping is above the Mott-transition. This cryogenic design with 2x base doping, nevertheless, improves freezeout in the intrinsic base and reduces  $R_B$  at 43 K.

Note that the modification to the control process flow only occurs in the SiGe



**Figure 27:** The doping (in logarithm scale) and Ge (in linear scale) profiles for the control and optimized SiGe HBTs.

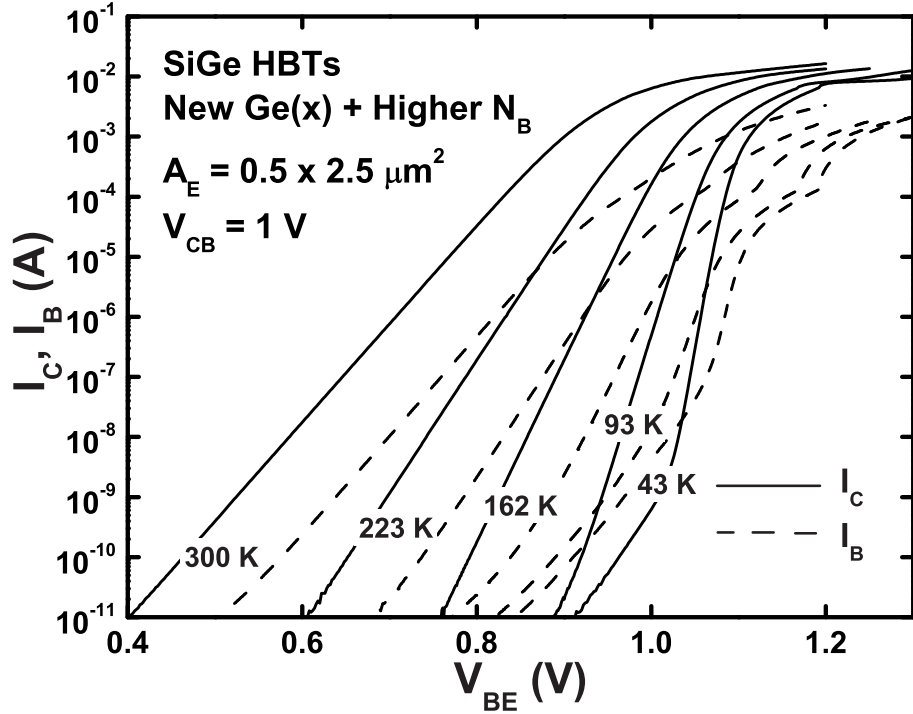
epitaxial base deposition and does not affect the CMOS transistors or other passives on the same wafer. Nor does it add cost to fabrication, simply an alternate epi base deposition. A slight shift in collector doping is expected due to the inevitable adjustment of the buffer layer thickness in order to keep the same collector profile among different designs.

### 3.3 Measurement Results

SiGe HBTs with three different profiles were fabricated and tested at 43 K to 300 K, including the “*control*” devices, the SiGe HBTs with the “*new Ge(x)*” profile, and the transistors with both the new Ge(x) and the increased base doping (“*new Ge(x)*”).



+ higher  $N_B$ "). The cryogenic DC measurements were made using a closed-cycle cryostat while the AC data were taken in a custom on-wafer, open-cycle, liquid helium cryogenic probe station. An Agilent 4156 Parameter Analyzer and an Agilent E8364B PNA Network Analyzer were used for measurements. All measured devices have emitter areas of  $0.5 \times 2.5 \mu\text{m}^2$  for comparisons. Fig. 28 shows the near-ideal Gummel characteristics obtained across temperature, indicating the feasibility of these devices for the entire temperature range. Transistor yield was not affected.

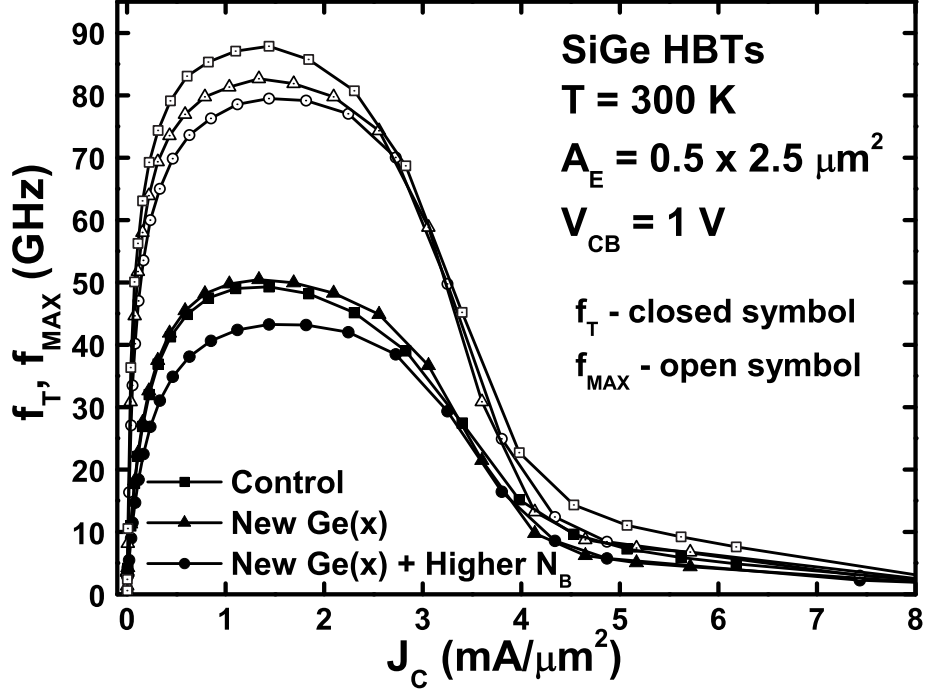


**Figure 28:** Forward Gummel characteristics of a SiGe HBT (*new Ge(x) + higher  $N_B$* ) at different temperatures.

### 3.3.1 RF Improvements

Figs. 29 and 30 show the measured  $f_T$  and  $f_{max}$  vs. collector current density ( $J_C$ ) at 300 K and 43 K. The *new Ge(x)* profile has little impact on  $f_T$  and  $f_{max}$  at 300 K,

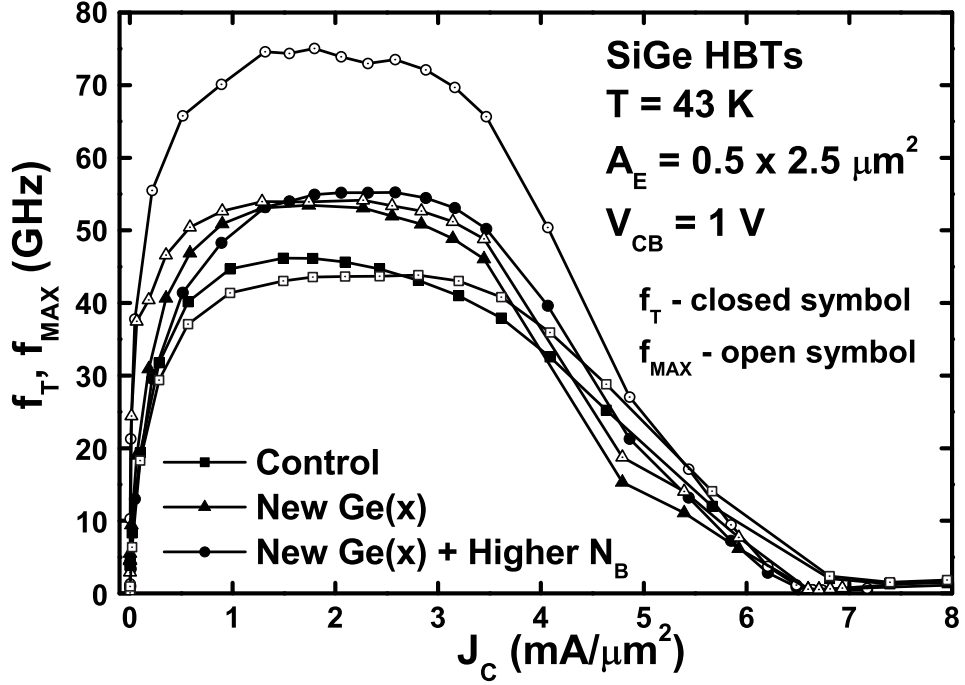
as expected; however, its advantage is clearly shown at 43 K, where the peak  $f_T$ 's are 46.2, 53.5, and 55.6 GHz for three profiles (*control*, *new Ge(x)* and *new Ge(x) + higher  $N_B$* ), respectively.



**Figure 29:** Measured  $f_T$  and  $f_{max}$  vs.  $J_C$  for the three profiles at 300 K.

Meanwhile, the higher base doping lowers the peak  $f_T / f_{max}$  from 49.3 / 87.8 GHz to 43.2 / 79.4 GHz at 300 K, because of the higher  $\tau_B$ , as expected. Base freezeout is partially mitigated with the new profile, however, as evidenced by the peak  $f_{max}$  at 43 K (43.8 GHz for the *control*, 53.9 GHz for *new Ge(x)*, and 73.5 GHz for *new Ge(x) + higher  $N_B$* ).

The peak  $f_T$  and  $f_{max}$  of all the three profiles across temperature are shown in Fig. (31). The  $f_{max}$  of a control device begins to degrade at 162 K when carrier freezeout onsets, and this effect is greatly reduced in the *new Ge(x) + higher  $N_B$*



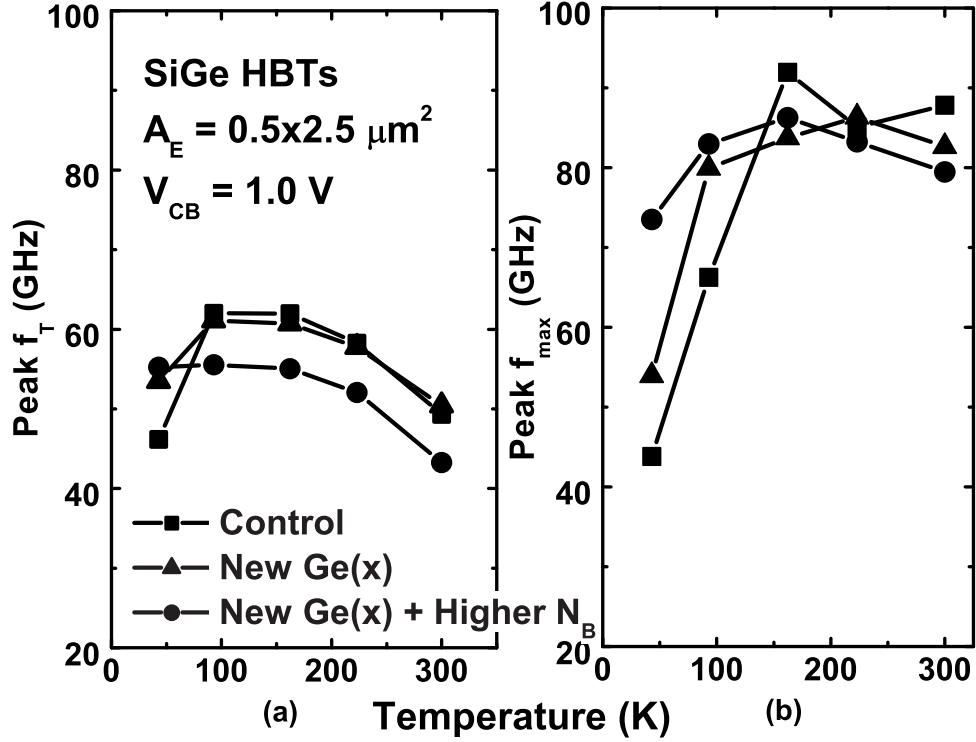
**Figure 30:** Measured  $f_T$  and  $f_{max}$  vs.  $J_C$  for the three profiles at 43 K.

profile. At 43 K, the *new Ge(x) + higher  $N_B$*  profile shows an overall improvement of 20% in  $f_T$  and 70% in  $f_{max}$  compared to the *control* device.

### 3.3.2 DC Characteristics

Fig. 32 and 33 show the DC current gain ( $\beta$ ) vs.  $I_C$  at 300 K and 43 K, respectively. In each figure, (a) shows the low injection operation with  $I_C$  on a logarithm scale and (b) shows the medium-to-high injection operation with  $I_C$  on a linear scale. The new Ge profile has little impact on  $\beta$  at 300 K, as can be seen by comparing the *control* and *new Ge(x)* profiles.

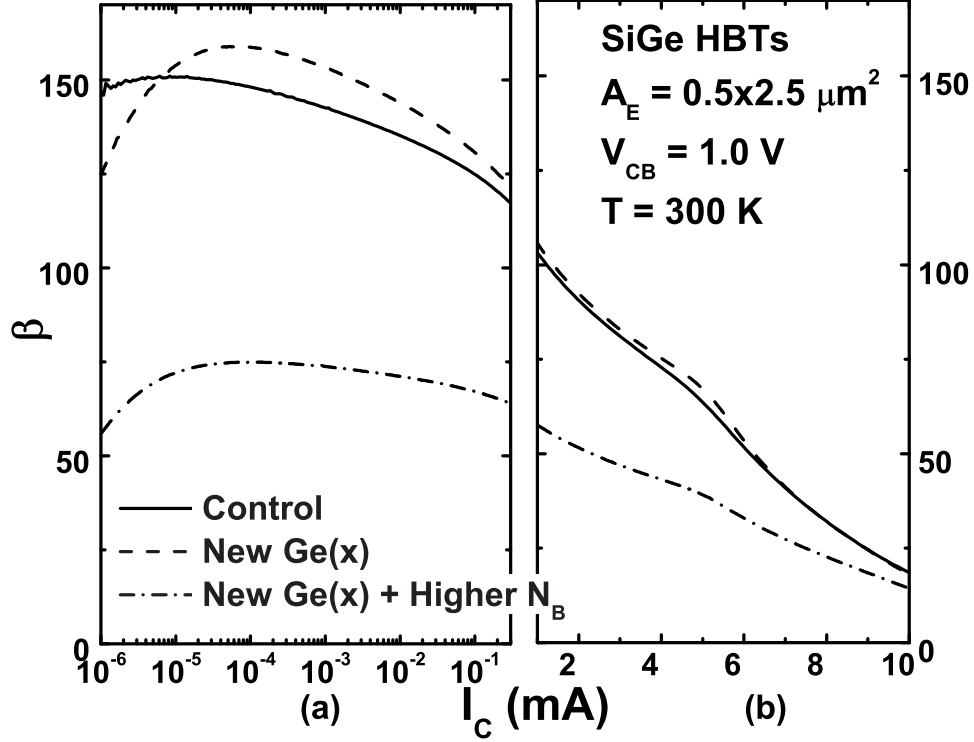
Doubling the base doping increases base Gummel number and thus reduces  $\beta$  throughout the entire  $I_C$  region in the *new Ge(x) + higher  $N_B$*  profile, at all temperatures. Additionally, the higher base doping causes the EB metallurgical junction



**Figure 31:** Peak  $f_T$  and  $f_{max}$  vs. temperature for the three profiles.

to shift closer to the emitter, reducing the Ge fraction at the junction, leading to a degraded  $\beta$  at low-injection. Interestingly, however, the *new Ge(x)* and *new Ge(x) + high  $N_B$*  profiles have similar  $\beta$  at 43 K at high injection (Fig. 33).

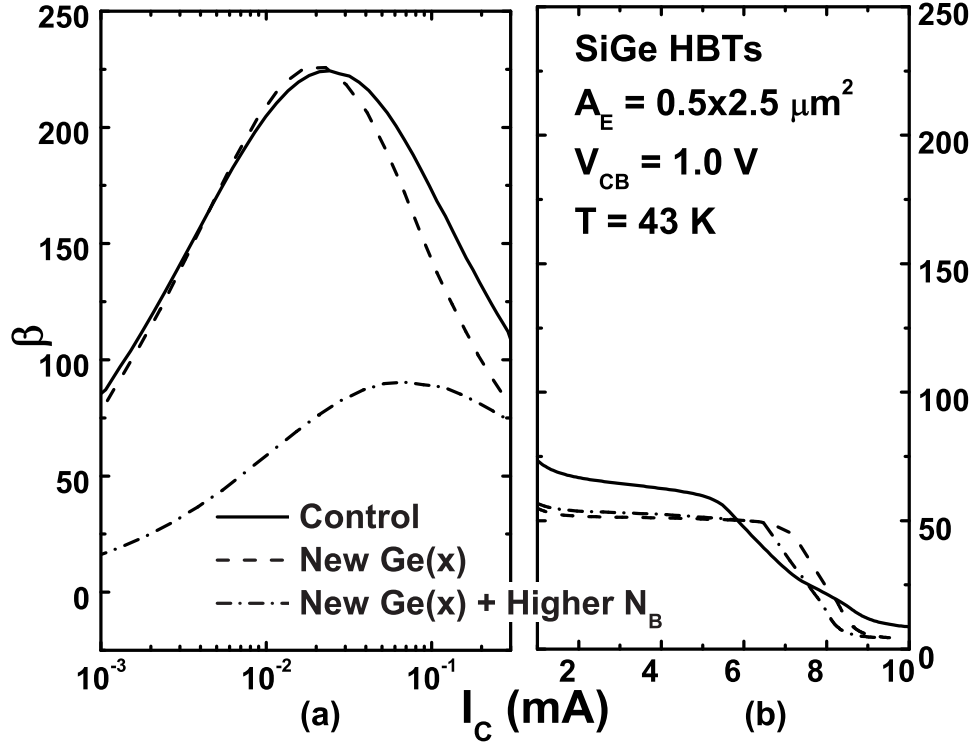
The peak  $\beta$  as functions of temperature for all three of the SiGe HBTs are plotted in Fig. 34. The *control* and *new Ge(x)* profiles have a similar temperature dependence for peak  $\beta$  because the only difference between these two lies in the Ge profile that extends to the CB depletion region, thus only affecting medium-to-high injection. The *new Ge(x) + higher  $N_B$*  profile deviates the *control* and *new Ge(x)* profiles because of the lower base doping and the lower Ge fraction in the EB junction. However, having a  $\beta$  between 50 and 100 from 43 to 300 K is more than adequate for the targeted RF circuit designs.



**Figure 32:** Measured  $\beta$  vs.  $I_C$  for the three profiles at 300 K.

### 3.4 Physical Origins

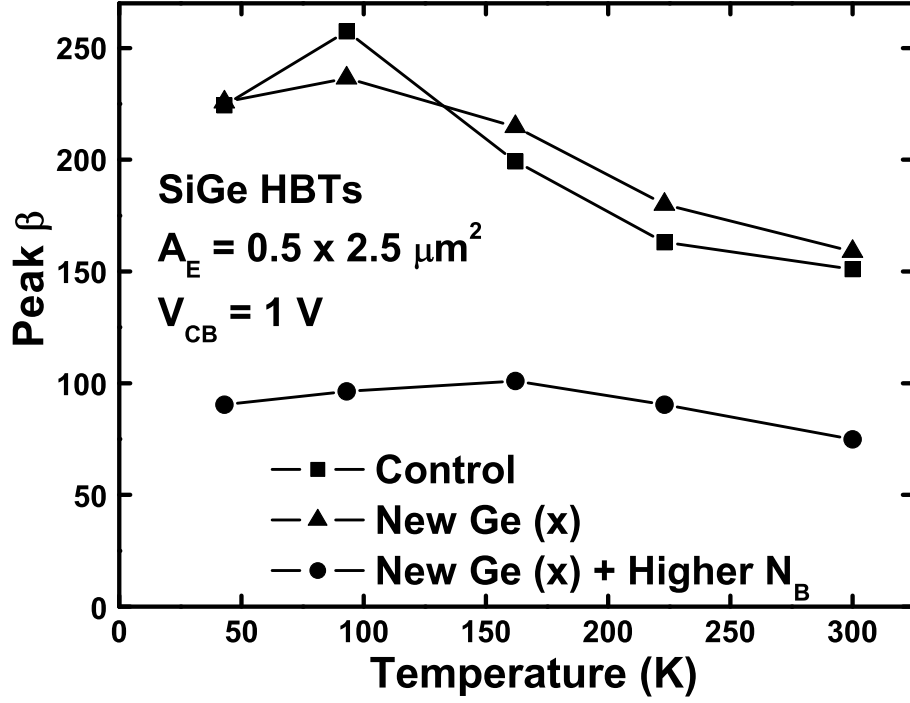
In order to shed light on the physics of how the new germanium profiles affect the high injection performance of SiGe HBTs at cryogenic temperatures, the transconductance ( $g_m$ ) as a function of  $I_C$  for all the three SiGe HBTs at 43 K and 300 K is plotted in Fig. 35. At low injection,  $g_m$  increases linearly with increasing  $I_C$ . This trend, nonetheless, tends to saturate at medium injection as Kirk effect and the voltage drop on the series resistance start to engage; beyond this point,  $g_m$  reaches its minimum as HBE begins to dominate. This  $g_m$  vs.  $I_C$  behavior at room temperature is mainly modulated by the collector doping, and hence there is hardly any difference between the three profiles at 300 K (dashed curves in Fig. 35).



**Figure 33:** Measured  $\beta$  vs.  $I_C$  for the three profiles at 43 K.

HBE at low temperatures, however, plays a much more important role, as  $g_m$  continues increasing until it triggers the positive feedback mechanism as described in the previous chapter, when it increases precipitously (shown by the solid lines in Fig. 35). The *new Ge(x)* profile shows this effect at much higher currents at 43 K than the control profile, clearly indicating the effectiveness of the new Ge profile in delaying HBE at low temperatures. Note that the effective collector doping is a little higher in the *new Ge(x) + higher  $N_B$*  profile, leading to a slightly higher  $I_C$  needed to engage HBE.

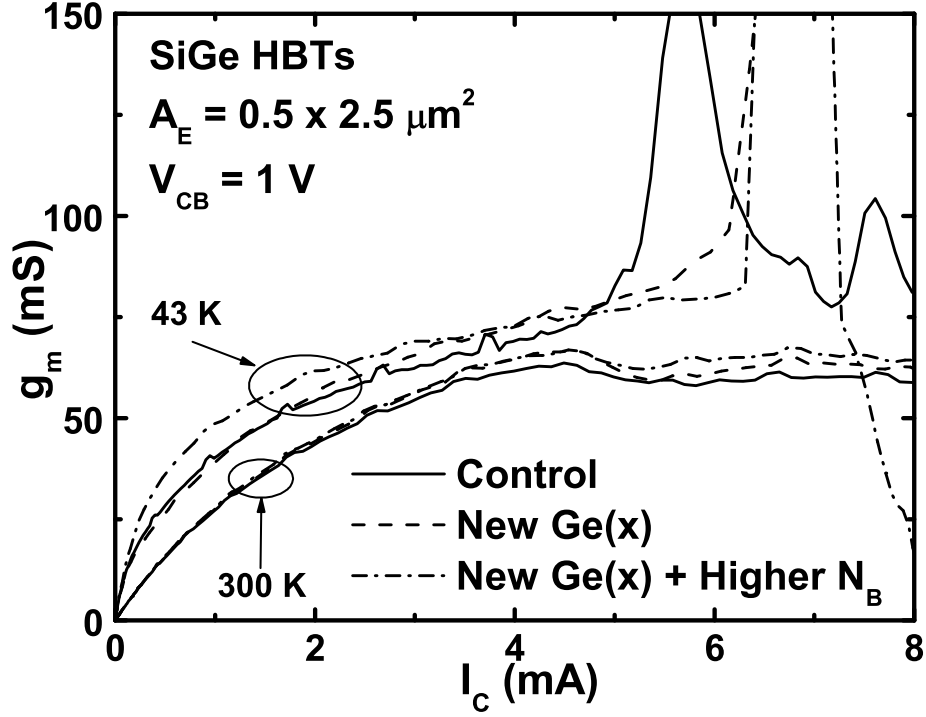
The overall  $g_m$  vs. temperature at  $I_C=1.8$  mA, which is about the bias current for obtaining peak  $f_T$ , is shown in Fig. 36. By delaying HBE to a higher  $I_C$ , the *new Ge(x)* profile greatly enhances  $g_m$  at high injection levels, resulting in an enhanced



**Figure 34:** Peak  $\beta$  vs. temperature for the three profiles.

$f_T$  even with the same impurity doping profiles.

The  $R_B$  as a function of  $I_C$  is extracted at 43 K and 300 K in Fig. 37, using the impedance circle method. The  $R_B$  of the *new Ge(x) + high  $N_B$*  profile is 15% lower than that of the control profile due to the lower  $R_{bi}$  resulting from the higher base doping. The difference in  $R_B$  is much more substantial at 43 K, as  $R_B$  nearly triples at low injection for the *control* and *new Ge(x)* profiles because of carrier freezeout, whose effects can be greatly suppressed with a higher base doping, as shown by the *new Ge(x) + higher  $N_B$*  profile data at 43 K. More importantly, the steep  $R_B$  vs.  $I_C$  at 43 K is the combined result of carrier freezeout and base push-out, as modeled in [149], and comes together with higher  $C_{CB}$ , which can only be suppressed with higher base doping. This is evidenced by comparing the *control* and *new Ge(x) +*

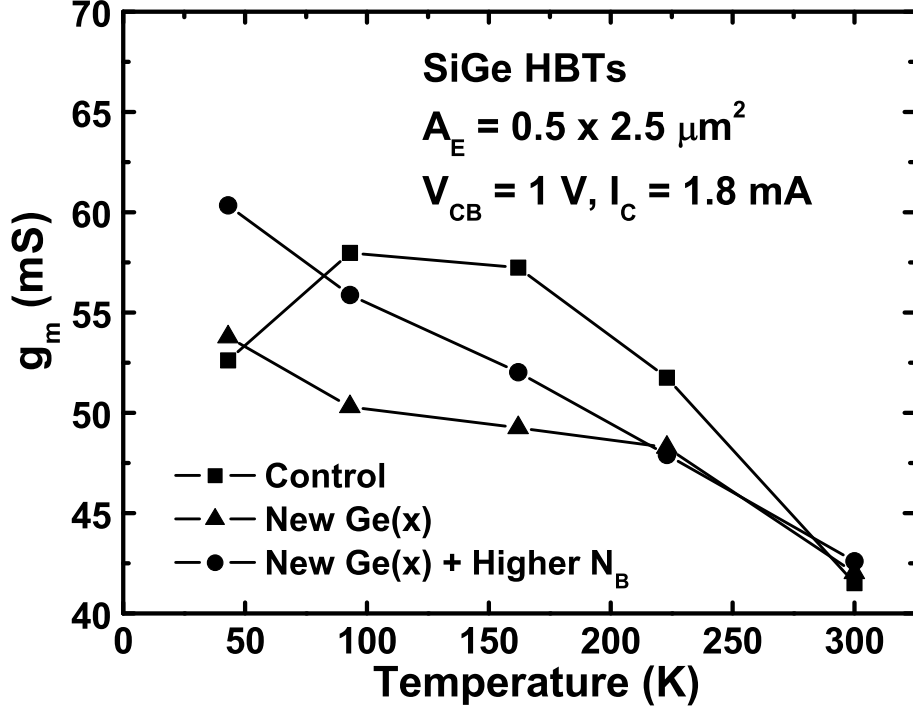


**Figure 35:** Measured  $g_m$  vs.  $I_C$  for the three profiles at 43 K and 300 K.

*higher*  $N_B$  profiles at  $I_C=1.8$  mA at 43 K, where  $f_T$ ,  $R_B$ , and  $f_{max}$  improve by 20%, 15%, and 70% in the optimized profile. In summary, suppressing carrier freezeout is crucial in obtaining high  $f_{max}$  at 43 K at this technology node.

The  $BV_{CEO}$  and  $BV_{CBO}$  of all the three profiles were measured across temperature and plotted in Fig. 38. The  $M-1$  generally increases with cooling, as expected, leading to a  $BV_{CBO}$  1 V lower at 43 K than at 300 K. The *new Ge(x)* profile improves  $BV_{CBO}$  by changing the electric field in the CB depletion region, thereby offsetting the negative impact of the higher doping-induced  $BV_{CBO}$  degradation (thus the *control* and *new Ge(x) + higher  $N_B$*  profiles have similar  $BV_{CBO}$  across temperatures). Meanwhile, the lower  $\beta$  in the *new Ge(x) + higher  $N_B$*  profile improves  $BV_{CEO}$ , as shown in Fig. 38 (a). We point out, however, that despite the differences, all three



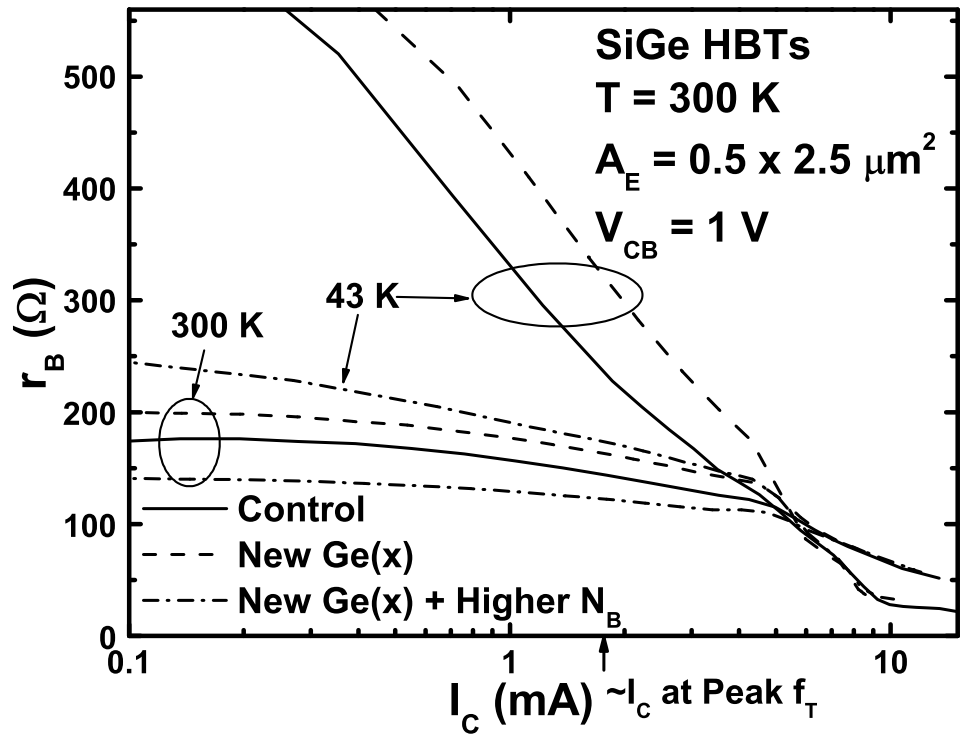


**Figure 36:** Measured  $g_m$  at  $I_C = 1.8$  mA vs. temperature for the three profiles.

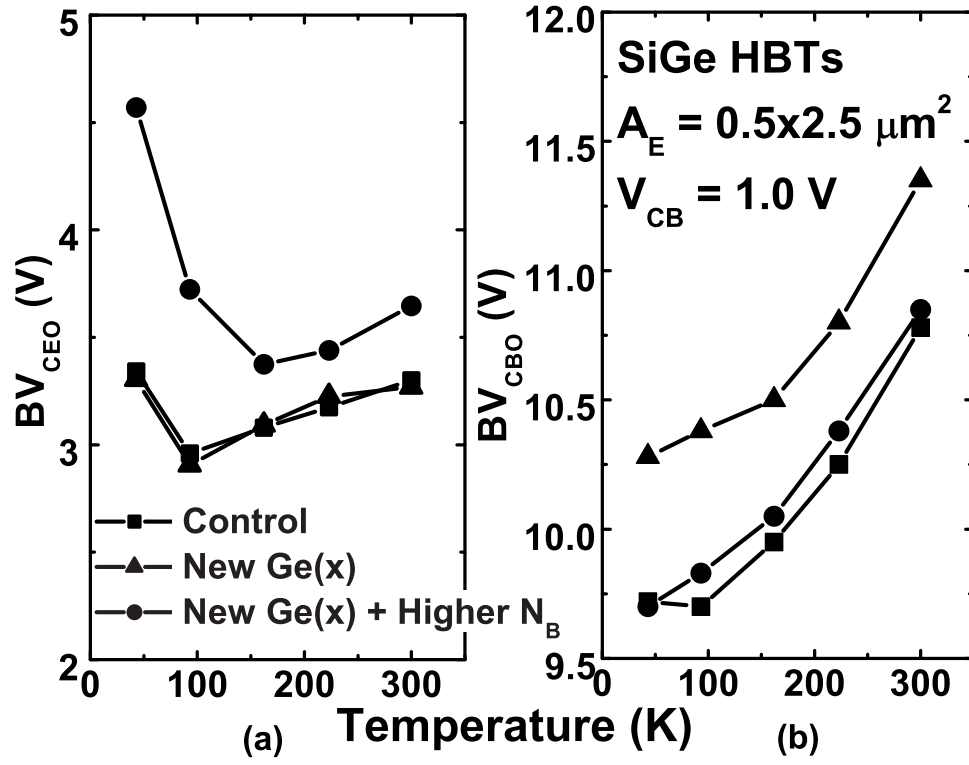
profiles have useful  $BV_{CEO}$  and  $BV_{CBO}$  values for real RF circuit applications across the entire temperature range.

### 3.5 Summary

We have demonstrated improved Ge and base doping profiles that optimize the cryogenic performance of the first-generation SiGe HBTs. Careful DC and AC analysis shows that the new profiles are effective in suppressing carrier freezeout and the heterojunction barrier effect at cryogenic temperatures, leading to a 20% higher  $f_T$  and a 70% higher  $f_{max}$  compared to the control SiGe HBT at 43 K.



**Figure 37:** Extracted  $R_B$  vs.  $I_C$  for the three profiles at 300 K and 43 K.



**Figure 38:** Measured  $BV_{CEO}$  and  $BV_{CBO}$  vs. temperature for the three profiles.

## CHAPTER IV

### PERFORMANCE LIMITS

The goal of achieving terahertz transistors within the silicon material system has generated significant recent interest. In this chapter we use operating temperature as an effective way of gaining a better understanding of the performance limits of SiGe HBTs and their ultimate capabilities for achieving terahertz speeds. Different approaches for vertical profile scaling and reduction of parasitics are addressed, and three prototype fourth-generation SiGe HBTs are compared and evaluated down to deep cryogenic temperatures, using both DC and AC measurements. A record peak  $f_T / f_{max}$  of 463 / 618 GHz was achieved at 4.5 K using 130 nm lithography (309 / 343 GHz at 300 K), demonstrating the feasibility of reaching half-terahertz  $f_T$  and  $f_{max}$  simultaneously in a silicon-based transistor. The  $BV_{CEO}$  of this cooled SiGe HBT was 1.6 V at 4.5 K ( $BV_{CBO} = 5.6$  V), yielding a record  $f_T \times BV_{CEO}$  product of 750 GHz-V (510 GHz-V at 300 K). These remarkable levels of transistor performance and the associated interesting device physics observed at cryogenic temperatures in these devices provide important insights into further device scaling for THz speeds at room temperature.

We also present a measured CML ring oscillator gate delay of 2.3 ps, a record for digital circuits in silicon-based technologies. In addition to higher cutoff frequency and lower collector-base capacitance, decreasing base resistance is also responsible for the higher switching speed at cryogenic temperatures. The self-heating characteristics of these SiGe HBT circuits are also investigated across temperature.

## 4.1 *Scaling and Cooling of SiGe HBTs*

Device scaling has been the principal driving force behind IC technology innovation for the past four decades. The vertical and lateral scaling of SiGe HBTs was reviewed in the first chapter. A different approach for improving performance is to cool the SiGe HBTs to reach their ultimate performance. Pioneering work on cooled SiGe HBTs demonstrated improved performance at cryogenic temperatures [39]. Such cooling-induced performance enhancements, nevertheless, were not dramatic, partly due to the relatively small bandgap grading used in first-generation SiGe HBTs, together with the finite base freezeout associated with the modest base doping [5]. However, the cryogenic operation of more aggressively-scaled modern SiGe HBTs is in many ways more equivalent to room temperature vertical and lateral scaling, as will be shown, and hence cooling has recently become a convenient and effective way of gaining a better understanding of the ultimate performance of SiGe HBTs at the limits of scaling.

### 4.1.1 **Bandgap Engineering and Vertical Scaling**

Bandgap engineering resulting from the presence of Ge in the base region of a SiGe HBT has many similarities to the desired results of classical vertical profile scaling of bipolar transistors. The Ge content at the base-emitter heterojunction side of the neutral base lowers the conduction-band barrier by  $\Delta E_{g,Ge}(0)$  at a fixed voltage bias and hence enhances electron injection. Combined with the Ge grading across the base ( $\Delta E_{g,Ge}(grade)$ ),  $\beta$  is greatly increased compared to a control Si BJT with the same doping profile. Ge-induced bandgap engineering also improves the AC performance of the device by reducing base and emitter transit times ( $\tau_E$ ). One needs to increase  $\Delta E_{g,Ge}(grade)/kT$  and  $\Delta E_{g,Ge}(0)/kT$  in order to improve  $\beta$ ,  $\tau_E$ , and  $\tau_B$ . An improved Early voltage,  $V_A$ , is also one of the benefits of bandgap engineering, leading to higher output resistance [39].

There are clearly two ways to achieve enhancement of all these factors: 1) vertical profile scaling, which allows, for fixed film stability, an increased  $\Delta E_{g,Ge}(0)$  and  $\Delta E_{g,Ge}(grade)$ ; or 2) decrease the operating temperature. Both methods lead to higher  $\beta$  and  $f_T$ .

#### 4.1.2 Reduction of Parasitics and Charging Times

Cooling can effectively reduce the primary performance-limiting parasitics of SiGe HBTs such as  $C_{BC}$  and  $R_B$ , while adjusting  $\tau_{BE}$ , without physically scaling the lateral dimensions (i.e., at fixed lithography node). This positive benefit of cooling can be explained from fundamental physics [150]. The Fermi-level ( $E_F$ ) of  $n$ -type non-degenerate Si in the collector  $E_F$  moves closer to the conduction band with cooling [150]. The consequent built-in potential ( $V_{bi}$ ) of the PN junctions is thus increased at cryogenic temperatures. The resultant wider space charge region at fixed bias thus gives rise to a decrease in  $C_{BC}$  at cryogenic temperatures. Moreover, the widening of  $W_{CSCL}$  at low temperatures does not increase  $\tau_{CSCL}$  because  $v_{sat}$  is also higher at low temperatures, which enables improved current-drive.

The increased base doping level clearly must increase with generational scaling and this plays an important role in base resistance at low temperatures. The carrier transport properties of Si (or SiGe) are very dependent on the doping level. As described in the first chapter, the associated carrier freezeout in Si (or SiGe) doped below  $10^{18}$  atoms/cm<sup>3</sup> at low temperature is detrimental to the SiGe HBT's cryogenic performance (e.g., to  $\beta$ ,  $R_B$ ,  $f_{max}$ , and noise figure). Increasing the base doping can gradually degenerate silicon from a semiconducting to a quasi-metallic behavior over temperature. Above the Mott-transition, all dopants remain active to deep cryogenic temperatures, and the resistivity is a monotonically increasing function of temperature [150]. Indeed, one can infer that  $R_B$  is significantly reduced with decreasing temperature in  $0.12 \times 10 \mu m^2$  third-generation SiGe HBTs, where the

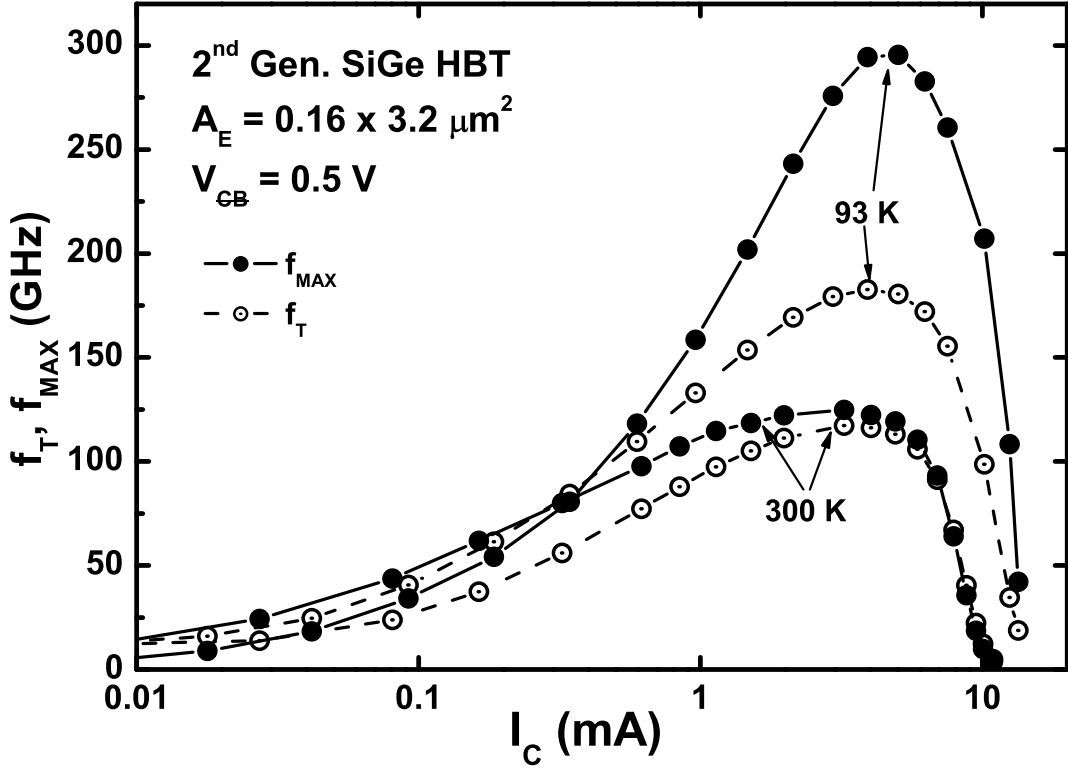
small-signal  $R_B$  equals 17.5, 18, 19, 22, and 32  $\Omega$  at 85, 120, 150, 200, and 300 K, respectively [7]. As discussed above, the direct benefit of reducing device parasitics is a higher  $f_{max}$ .

In addition,  $\tau_{BE}$  is very sensitive to the ambient temperature. Since both the emitter and the base in an advanced SiGe HBT are heavily doped (degenerate), the Fermi levels are usually considered to be at the edges of the conduction (for emitter) and valence (for base) bands, and  $V_{bi}$  of the BE junction at zero bias does not vary significantly across temperature. However, since higher  $V_{BE}$  is needed to operate the transistor at cryogenic temperature resulting from the lower intrinsic carrier concentration ( $n_i$ ), the BE junction width is effectively reduced during operation, increasing  $C_{BE}$  and decreasing the electron transit time through the BE SCR which changes  $\tau_{BE}$ . As discussed in Chapter 1, because of the difficulty in decoupling  $\tau_{BE}$  from  $\tau_B$  and  $\tau_E$ , cooling becomes a useful tool in tuning the depletion region width of the BE junction to achieve an optimized  $\tau_{BE}$ .

Fig. 39 shows the  $f_T$  and  $f_{max}$  of a second-generation SiGe HBT at 300 and 93 K. The  $f_T / f_{max}$  increase from 117 / 124 GHz at 300 K to 182 / 295 GHz at 93 K. Observe that the performance of this cooled second-generation SiGe HBT is comparable to the room-temperature performance of a (scaled) third-generation SiGe HBT featuring  $f_T / f_{max} = 207 / 285$  GHz at 300 K. Similarly,  $f_T / f_{max} = 260 / 310$  GHz can be achieved by cooling this third-generation SiGe HBT to 85 K [7], and this performance level is close to that of a fourth-generation SiGe HBT. Therefore, it is logical to explore the ultimate speed limits of SiGe HBTs by investigating the cryogenic performance of the state-of-the-art fourth-generation SiGe HBTs.

#### 4.1.3 Noise

Higher  $\beta$  and  $f_T$ , together with reduced parasitics ( $R_B$  and  $C_{BC}$ ) mutually optimize the broadband noise performance of a SiGe HBT. Not only can noise figure

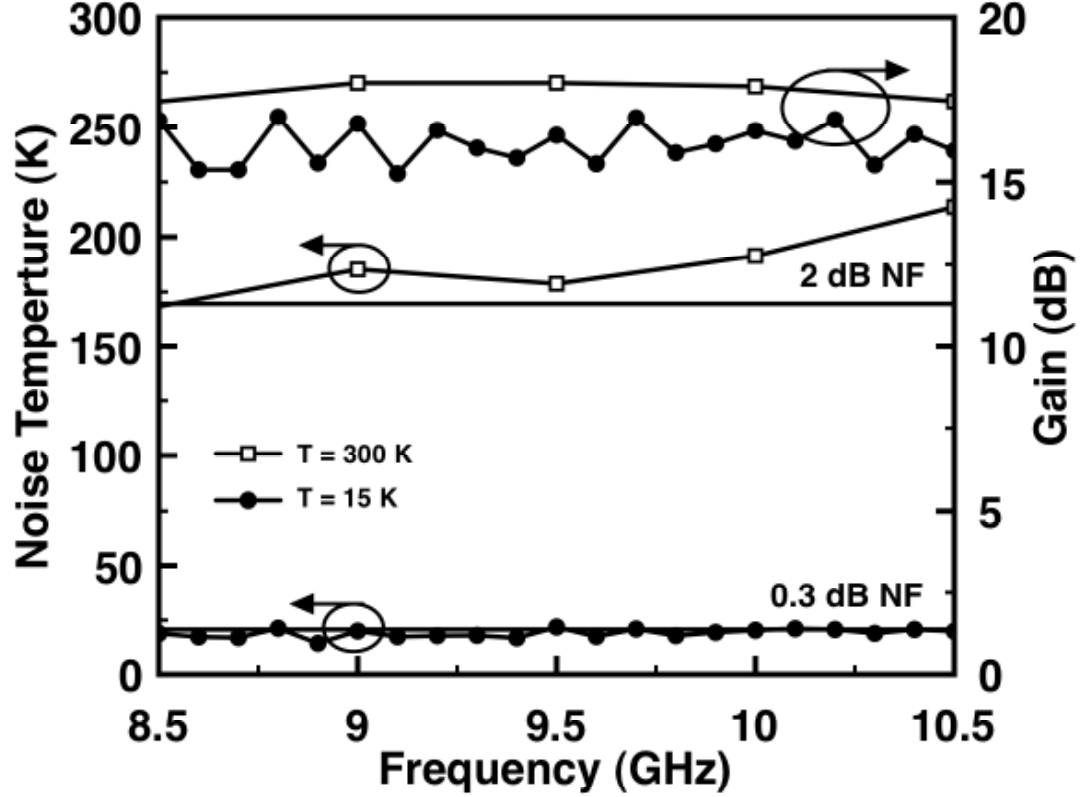


**Figure 39:**  $f_T$  and  $f_{max}$  as functions of collector current for a second-generation SiGe HBT at 300 K and 93 K.

be reduced by proper device scaling at 300 K, but it also improves naturally with cooling. At cryogenic temperatures, the shot noise is suppressed at fixed  $I_C$  because  $I_B$  is decreased via the higher  $\beta$ , the thermal noise directly depends on temperature, and  $R_B$  is reduced. In fact, a noise temperature of only 2 K in the 0.2–3-GHz range has been recently achieved in a third-generation SiGe HBT at 15 K [64].

To demonstrate that these remarkable levels of transistor noise performance can actually translate into useful circuits, we show very recent results on a cooled inductively-degenerated cascode SiGe HBT low noise amplifier (LNA) implemented in a third-generation SiGe technology, but which was designed for optimum room-temperature performance (2 dB noise figure and about 20 dB gain at 10 GHz at 300 K) [151]. Cooling this SiGe LNA produces a record sub-0.3 dB noise figure (below 20 K noise





**Figure 40:** Effective noise temperature (K) and gain (dB) of the X-band LNA at 300 K and 15 K (measured by T. Thrivikraman).

temperature) across X-band (8.5-10.5 GHz) at 15 K, as shown in Fig. 40 [29]. Though still noisier than the best InP HEMT cryogenic LNAs (with sub-0.1 dB noise figure at X-band [152]), cryogenically-operated SiGe amplifiers present some interesting application opportunities for the radio astronomy and remote sensing communities.

#### 4.2 Device Optimization for Half-THz Operation

The prototype fourth-generation SiGe HBTs used in this work were based on 130 nm third-generation device layouts (i.e., no substantial lateral scaling was used), and the basic device structure is described in [12]. They feature deep- and shallow-trench isolation, a boron-doped SiGe:C base layer with about 25% peak Ge concentration,

an SIC region, a reduced-thermal-cycle “raised-extrinsic-base” structure, and an in-situ phosphorus-doped polysilicon emitter.

Starting from the third-generation SiGe HBTs ( $f_T / f_{max}$ =207 / 285 GHz and  $BV_{CEO} / BV_{CBO}$ =1.8 / 5.5 V at 300 K), aggressive collector and base vertical scaling was performed, the emitter rapid thermal annealing (RTA) cycle was decreased, and the peak Ge in the base was increased to produce the prototype #1 of the fourth-generation SiGe HBTs. The goal was to achieve a high  $f_T$  by aggressive vertical scaling, and the resultant device reached an  $f_T$  of 350 GHz with an expected (degraded)  $f_{max}$  of about 200 GHz at 300 K. The breakdown voltage is also decreased due to the aggressive vertical scaling ( $BV_{CEO} / BV_{CBO}$ =1.4 / 5.0 V at 300 K). For a more balanced  $f_T$  and  $f_{max}$ , the device layout was changed from a CBE to a CBEBC structure (i.e., a double base and collector contact scheme) to reduce  $R_B$  and  $R_C$ , the base was further vertically-scaled but with a lower peak Ge content, and the collector scaling was more conservatively scaled, resulting in the prototype #2 device. In this prototype device, both  $f_T$  and  $f_{max}$  were balanced to a value above 300 GHz, at  $BV_{CEO} / BV_{CBO}$ =1.6 / 5.5 V (300 K). Finally, the base silicide to emitter spacing was decreased, and a significant portion of the crystalline section of the raised extrinsic base gets silicided to further reduce resistive parasitics to produce the prototype #3 SiGe HBT. Compared to prototype #1, prototype #3 employed a slightly more conservative vertical profile, while focusing on dramatically reducing lateral parasitics. Therefore, the  $f_{max}$  (350 GHz) and  $BV_{CEO} / BV_{CBO}$  (1.7 / 5.6 V) were designed to be higher (better) than for prototype #1, despite a minor reduction in  $f_T$  (300 GHz).

#### 4.2.1 DC Characteristics

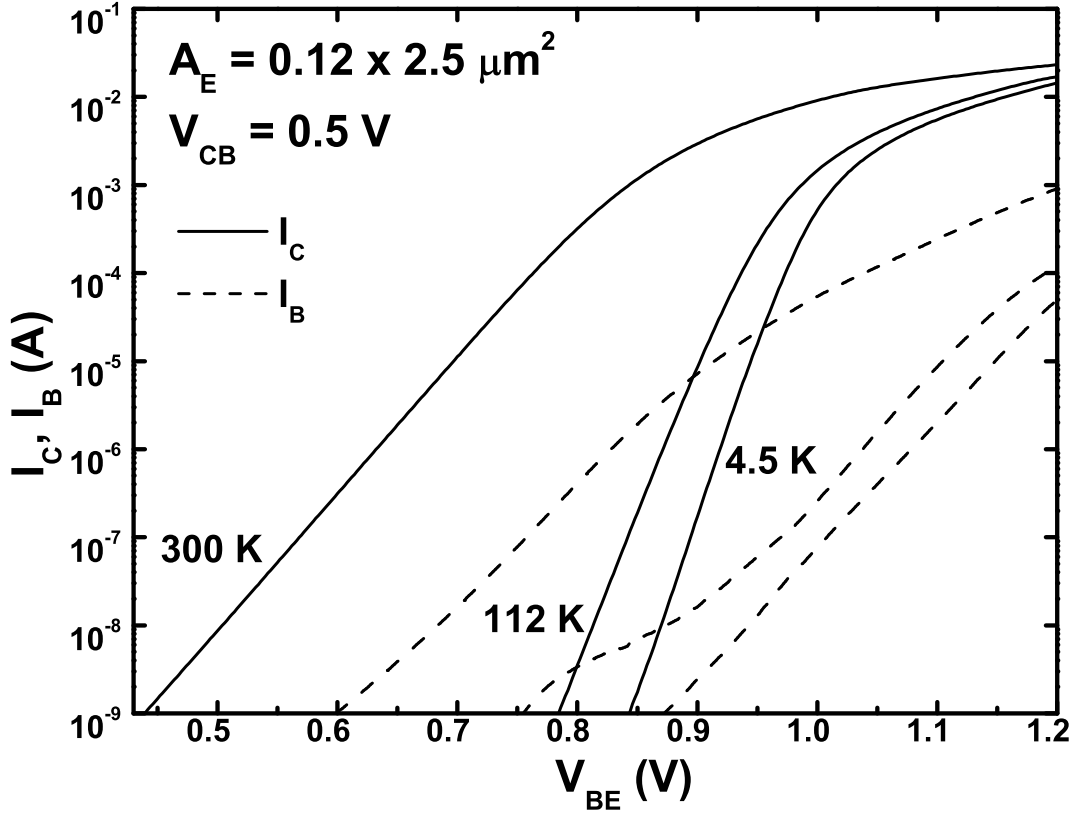
The setup for the cryogenic measurements (both DC and AC) were described in [43] and [23]. The cryogenic measurements were made using a custom on-wafer, open-cycle, liquid helium probe system. The thermometry was calibrated by comparing

the device characteristics inside the system with one measured directly immersed in liquid nitrogen (77.3 K). The accuracy of the thermometry is believed to be better than 1 K. The lowest chuck temperature was independently verified to be 4.5 K (liquid helium = 4.2 K), although we believe the actual device temperature is at least several kelvins above this temperature due to the inherent thermal loading of the probes when in contact with the wafer. For consistency the chuck temperature is specified in our data. The measured data are repeatable and have been exhaustively verified using a variety of techniques.

Exhaustive DC measurements in the temperature range of 4.5 – 300 K were made on prototype #3 SiGe HBTs with emitter areas of  $0.12 \times 2.5 \mu m^2$ , all showing reasonably “clean” Gummel characteristics at 300, 112, and 4.5 K (Fig. 41). The peak transconductance ( $g_m$ ) monotonically increases from 72.3 mS ( $241 \text{ mS}/\mu m^2$ ) at 300 K to 113 mS ( $377.5 \text{ mS}/\mu m^2$ ) at 4.5 K.

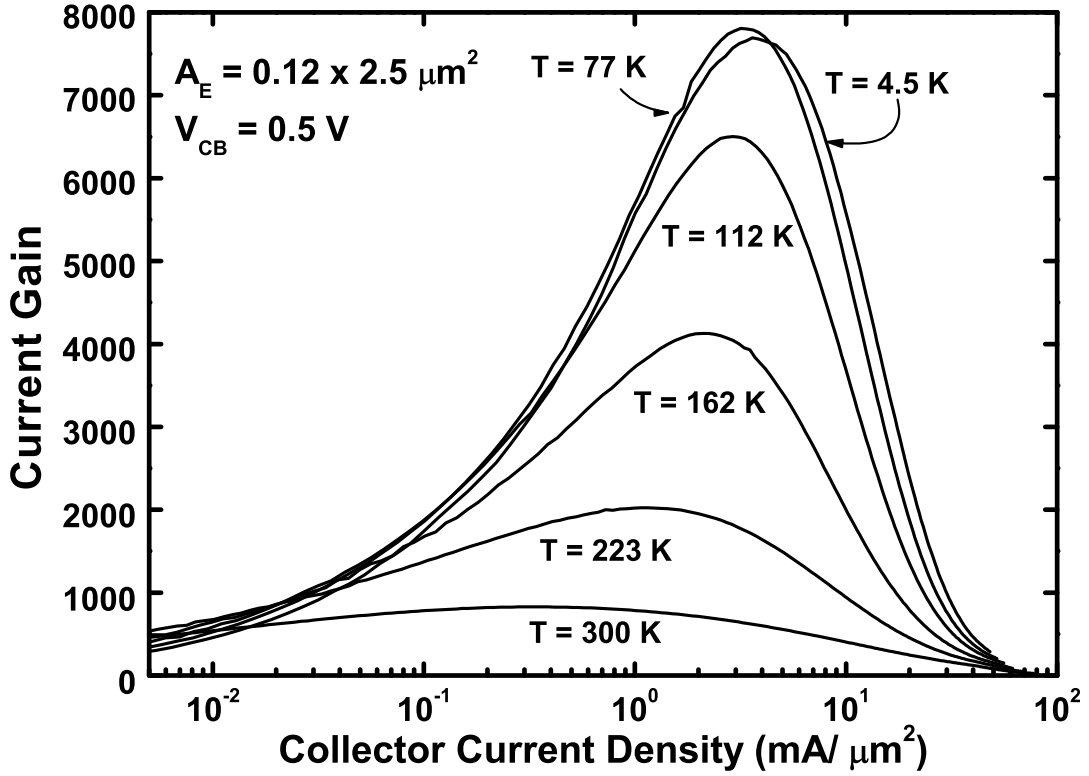
Current gain ( $\beta$ ) increases with cooling (Fig. 42), as expected, due to the Ge-induced band offset and the heavily doped base region that partially offsets the emitter-doping-induced bandgap narrowing. As the vertical profile scaling of this device is more aggressive than the third-generation HBT reported in [12], but more conservative than prototype #1, the peak  $\beta$  (7,900 at 77 K) is also between that of the previous two devices (3,800 at 85 K in third-generation and 9,297 at 77 K in prototype #1).

Forced- $I_B$  and forced- $V_{BE}$  output characteristics at 4.5 K are plotted in Fig. 43. The near-ideal families of output curves confirm the possibility of the cryogenic application of this device for real circuits. The negative differential resistance (NDR) effect under forced- $I_B$  input drive is related to modified tunneling/recombination processes [7]. This device is capable of sourcing more than  $40 \text{ mA}/\mu m^2$  of current density across the entire temperature range. The open-base breakdown voltage ( $BV_{CEO}$ ), measured from the open-base output characteristics and confirmed by the base-reversal points



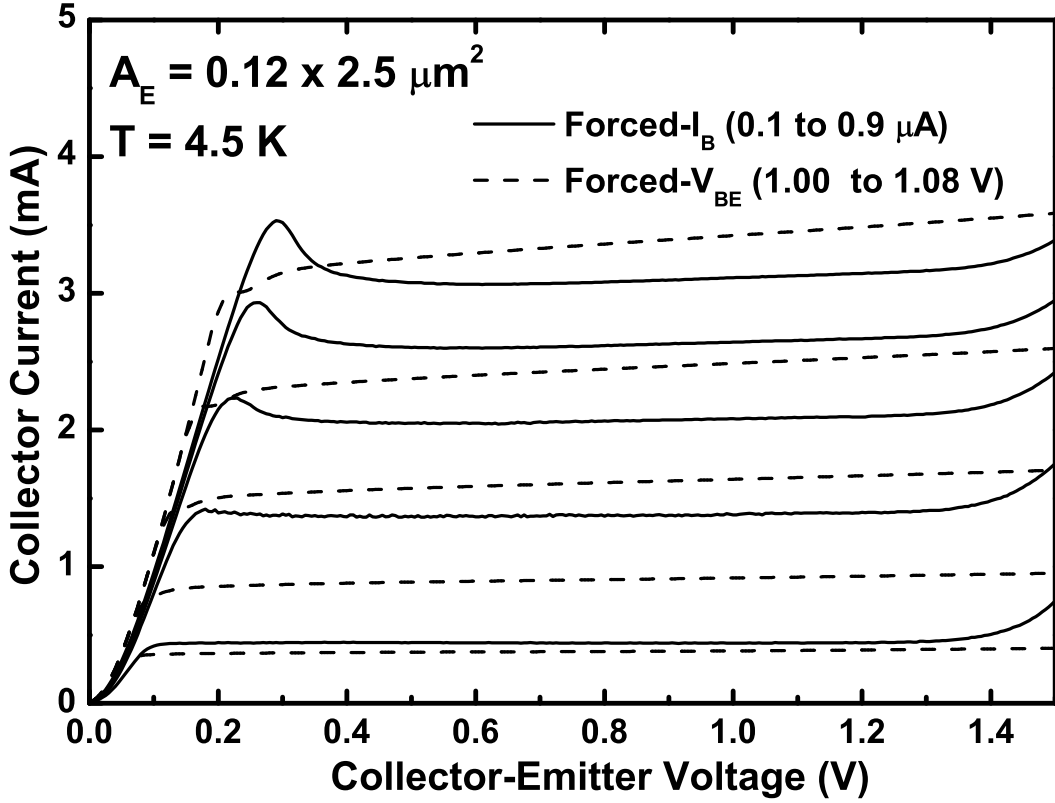
**Figure 41:** Forward Gummel characteristics for a fourth-generation prototype #3 SiGe HBT at 300, 112, and 4.5 K.

in a common-base, forced- $V_{BE}$  measurement [39], are 1.70, 1.63, and 1.62 V at 300, 112, and 4.5 K, respectively. The less aggressive vertical scaling indeed helps increase breakdown voltages across the entire temperature range. It is important to observe from Fig. 43 that  $BV_{CEO}$  at all  $I_C$  levels are always above 1.5 V, clearly of use in a circuits context. This is important in terms of safe operating area (SoA), and represents a major advantage of SiGe HBTs over III-V HBTs. The low  $\beta$  for cryogenically-operated SiGe HBTs at low injection improves  $BV_{CEO}$  by effectively reducing  $\beta(M - 1)$  product, as shown Fig. 42;  $BV_{CEO}$  at high injection depends on  $\beta$ , the  $M$  factor,  $R_C$ ,  $R_E$ , and  $V_{BE}$  that is increased at cryogenic temperatures due to the decrease of intrinsic carrier density.



**Figure 42:** Current gain at  $V_{CB} = 0.5$  V for a  $0.12 \times 2.5 \mu m^2$  SiGe HBT (fourth-generation prototype #3) as a function of collector current density, at 300, 223, 162, 112, 77, and 4.5 K.

In modern RF design,  $BV_{CBO}$  is often a more accurate upper limit for bias than  $BV_{CEO}$  [153], and DC bias beyond  $BV_{CEO}$  and even near  $BV_{CBO}$  can in fact be utilized in certain important classes of circuits (e.g., amplifiers) [154]. The measured  $BV_{CBO}$  of the present transistor remains at 5.6 V across temperature (4.5 K to 300 K), which bodes well for maintaining useful operating voltage with scaling for THz speeds. The constant  $BV_{CBO}$  over temperature implies a stronger Zener tunneling participation in those transistors than in slower transistors (first-, second-, and third-generations), as avalanching alone causes a positive temperature coefficient for the junction breakdown voltage.

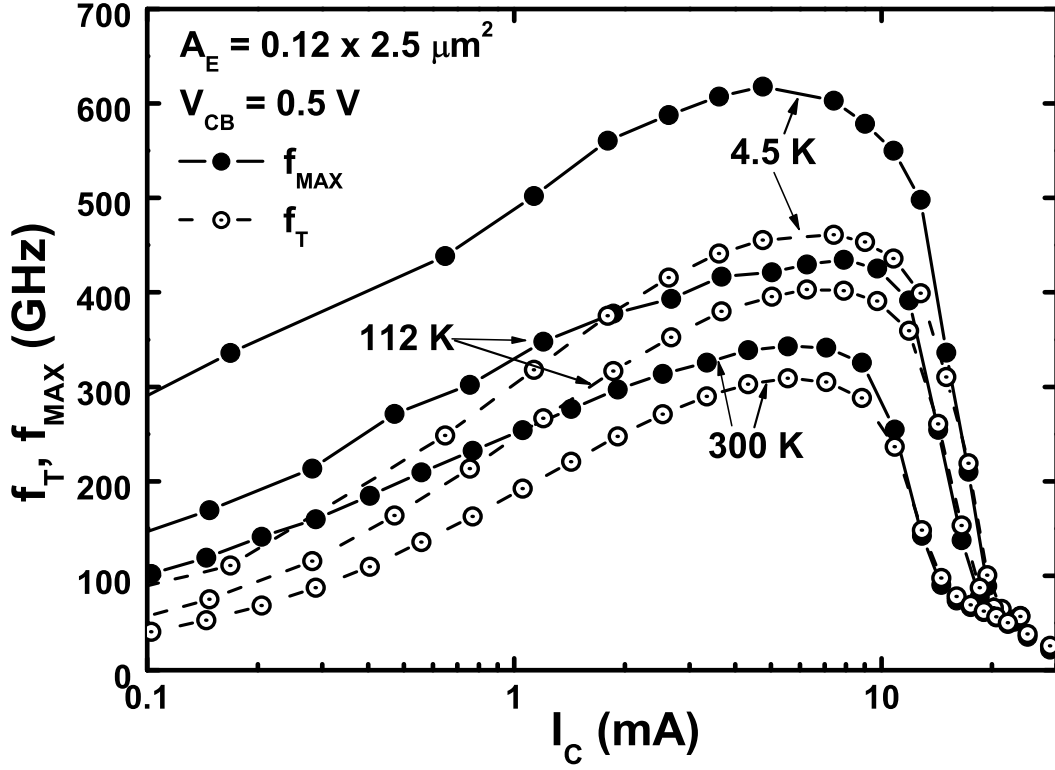


**Figure 43:** Forced- $I_B$  and forced- $V_{BE}$  output characteristics of a  $0.12 \times 2.5 \mu m^2$  SiGe HBT (fourth-generation prototype #3) at 4.5 K.

#### 4.2.2 AC Characteristics

Standard calibration and de-embedding techniques were used at each temperature, and the RF measurements on two different  $0.12 \times 2.5 \mu m^2$  SiGe HBTs were taken, showing similar results.

Fig. 44 shows the  $f_T$  and  $f_{max}$  of a SiGe HBT at 300, 112, and 4.5 K, extrapolated from  $h_{21}$  and  $U$  at 30 GHz, with the assumption that the gain rolls off at -20 dB/dec (Fig. 45). Clearly,  $f_{max} / f_T$  increase monotonically as the temperature decreases, growing from 343 / 309 GHz at 300 K to a record-high 618 / 463 GHz at 4.5 K. The dependence of the extracted  $f_T$  and  $f_{max}$  on extrapolation frequency is shown in Fig. 46, confirming the validity of the extraction method. The reported value of

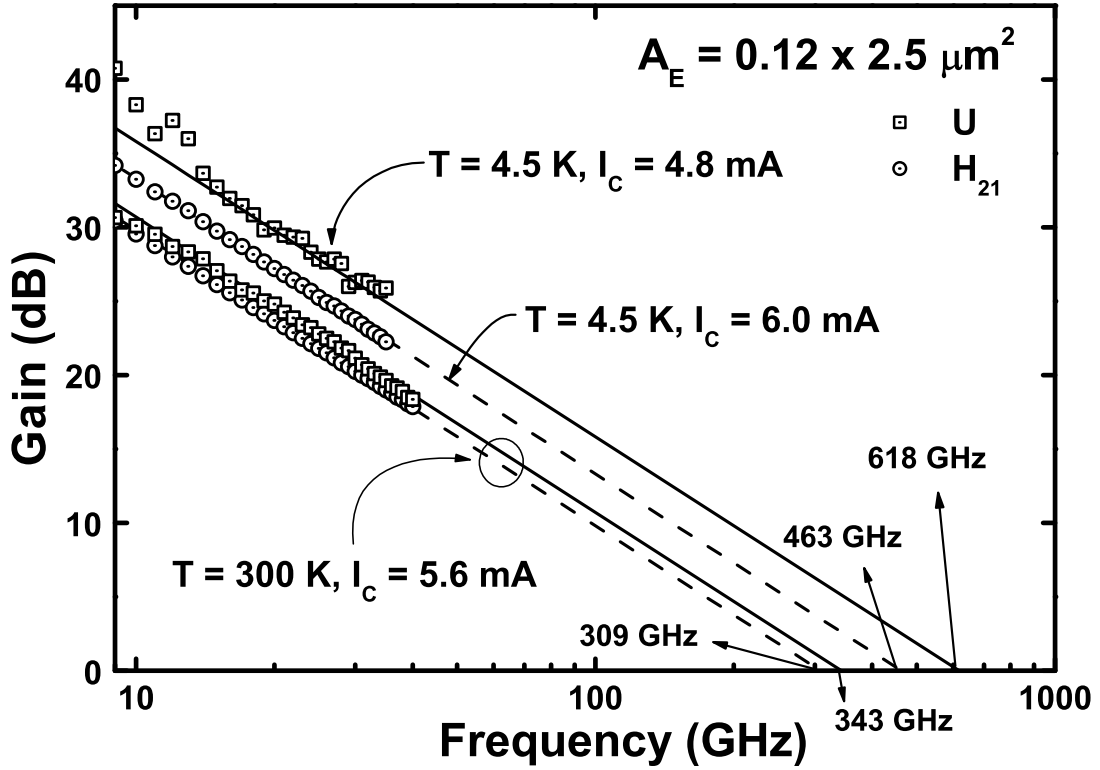


**Figure 44:**  $f_T$  and  $f_{max}$  as functions of collector current for a  $0.12 \times 2.5 \mu\text{m}^2$  SiGe HBT (fourth-generation prototype #3) at 300, 112, and 4.5 K.

618 GHz  $f_{max}$  is conservative.

Fig. 47 shows the measured S-parameters at 300, 112, and 4.5 K on a Smith chart for the frequency range of 3 GHz to 35 GHz, at peak  $f_{max}$  bias, showing the temperature dependence of the S-parameters. One can estimate the input and output matching points by conjugating these parameters at various frequency points from the Smith chart.

In Fig. 48, total delay ( $\tau_{EC} = 1/2\pi f_T$ ) versus reciprocal collector current density ( $1/J_C$ ) at 300, 112, and 4.5 K is shown. The extrapolated transit time ( $\tau_f$ ) decreases from 420 fs at 300 K to 330 fs at 112 K to 300 fs at 4.5 K. The transit time is slightly larger than prototype #1 (360 fs at 300 K and 270 fs at 4.5 K), mainly because of the increase in the base transit time across the slightly less aggressive base profile. The



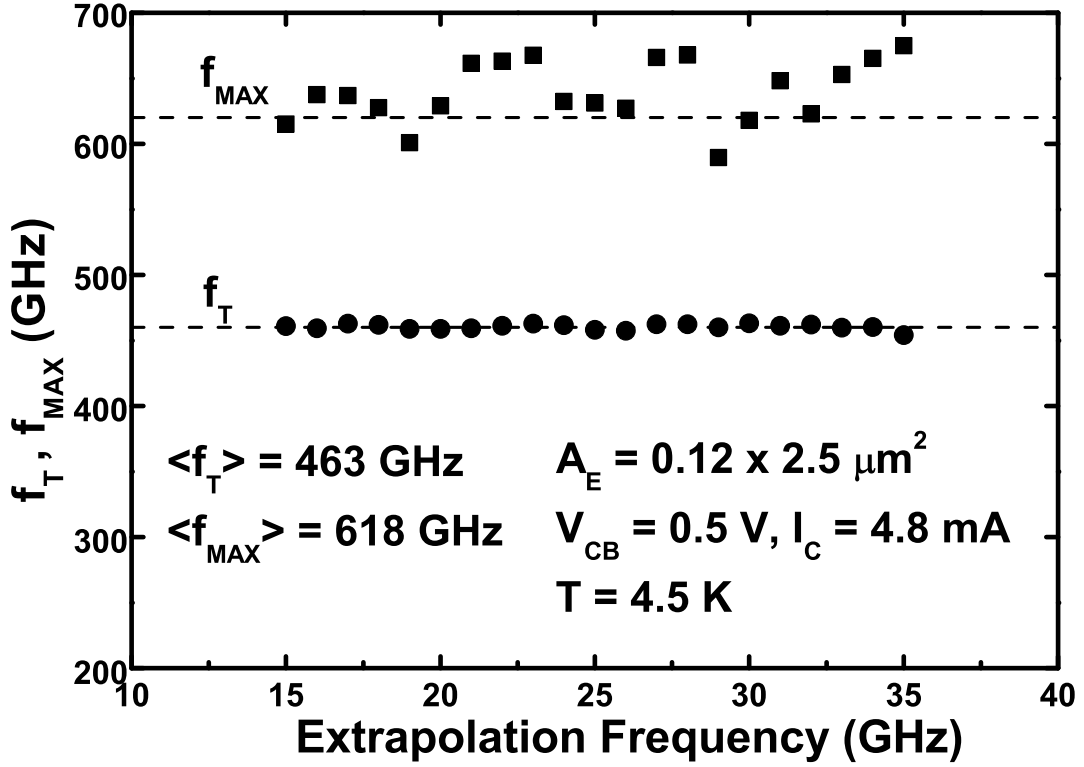
**Figure 45:** Small-signal power gain ( $U$ ) and current gain ( $h_{21}$ ) for a  $0.12 \times 2.5 \mu\text{m}^2$  SiGe HBT (fourth-generation prototype #3) at 300 and 4.5 K.

slope of the curve at high  $1/J_C$  is proportional to the total depletion capacitance, which decreases with cooling, mostly due to the increase in the CB junction built-in potential with cooling.

#### 4.2.3 Cryogenic Performance Limits

The key transistor DC and AC parameters of a  $0.12 \times 2.5 \mu\text{m}^2$  prototype #3 SiGe HBT across temperature are summarized in Table 3. One can easily observe the favorable impact of cooling on DC parameters ( $\beta$  and  $g_m$ ) and device speed ( $\tau_f$ ,  $f_T$ , and  $f_{max}$ ). As indicated above, this device maintains a relatively constant breakdown voltage ( $BV_{CEO}$  or  $BV_{CBO}$ ) across the entire temperature range. Insight for further device scaling can be obtained by comparing the measured results of the three different

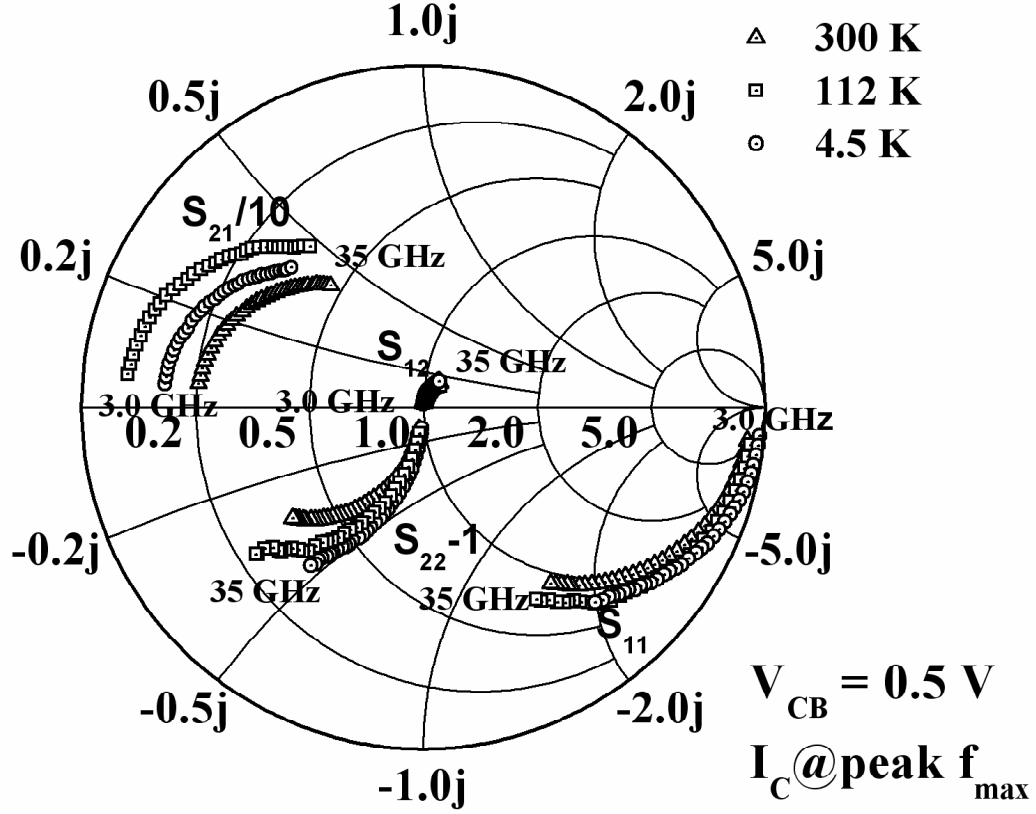




**Figure 46:**  $f_T$  and  $f_{max}$  extracted at various frequencies between 15 GHz and 35 GHz based on -20 dB/dec roll-off of  $h_{21}$  and  $U$ .

scaling prototypes.

The  $f_T$  and  $f_{max}$ , as functions of temperature, for all three prototype SiGe HBTs are compared in Fig. 49. While the peak  $f_T$  is mainly determined by the degree of “aggressiveness” in the vertical scaling and shows a similar trend in temperature dependence for all three prototypes, the slope of the peak  $f_{max}$  as a function of temperature is more associated with the different temperature dependencies of the various parasitics of the transistor. Reducing lateral parasitics (as in prototype #3) helps increase  $f_{max}$ , especially at cryogenic temperatures. To provide the context for further scaling, all of the present results are compared with the previously-reported 300 K  $f_T / f_{max}$  values from IBM and other companies in Fig. 50. An aggregate  $f_T + f_{max}$  above 1 THz can indeed be achieved in SiGe HBTs via careful device

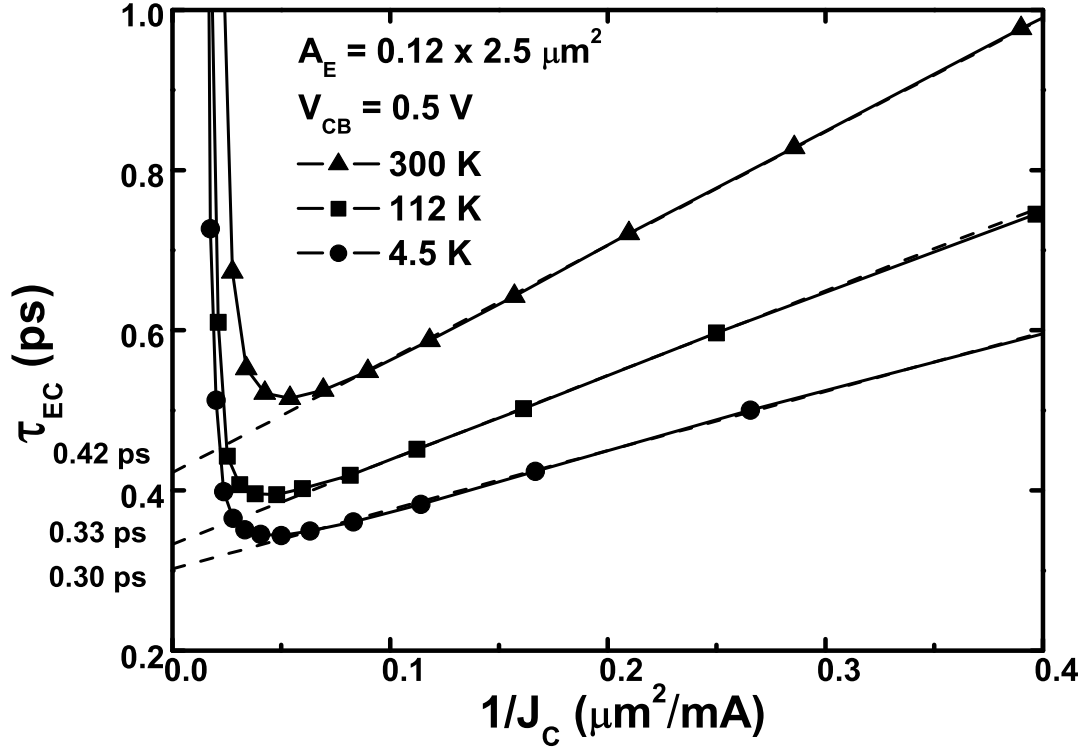


**Figure 47:** S-parameters ( $S_{11}$ ,  $S_{12}$ ,  $S_{21}/10$ , and  $S_{22} - 1$ ) at peak  $f_{max}$  for a  $0.12 \times 2.5 \mu m^2$  SiGe HBT (fourth-generation prototype #3) at 300, 112, and 4.5 K.

scaling and operation at cryogenic temperatures. Given the discussion above on the inherent similarities between cooling and further scaling, this leads us to believe that a similar aggregate 1 THz should in principle be achievable at 300 K.

### 4.3 Gate Delay of Cryogenic Ring Oscillators

The chips used in this section were fabricated in IHP's  $0.13 \mu m$  SiGe BiCMOS technology SG13 [155]. As shown in Fig. 51 (b), the SiGe module includes a SiGe:C base layer, shallow trench isolation, and a low-cost, deep-trench-free, implanted collector well process. The final rapid thermal annealing enables a vertical scaling that results in a 300 K  $f_T$  of 240 GHz, while the minimized CB junction and the lower  $R_B$  yielding a high  $f_{max}$  of 330 GHz at 300 K.



**Figure 48:** Total delay versus reciprocal collector current density at constant  $V_{CB}$  for a  $0.12 \times 2.5 \mu\text{m}^2$  SiGe HBT (fourth-generation prototype #3), at 300, 112, and 4.5 K.

The bipolar ring oscillators consist of 53-stage CML inverter circuits with differential inputs/outputs (Fig. 51 (a)). The tail current of each stage is controlled via a simple current mirror (Fig. 51 (c)). The value of current flowing through each stage is determined by the resistors  $R_L$ ,  $R_E$ , and an additional current tuning terminal. The output buffer is similar to each ring oscillator stage, and the single-ended peak-to-peak  $\Delta V$  is 300 mV. For each ring oscillator, the proper  $R_L$  was chosen to maintain a constant  $\Delta V = 300 \text{ mV}$  at different levels of current per stage ( $I_{gate}$ ).

#### 4.3.1 Cryogenic Ring Oscillator Performance

The CML ring oscillators were measured from 4.5 K to 294 K on a Lakeshore CPX-HF cryogenic probe station. Measured minimum CML ring oscillator gate as a

**Table 3:** Key parameters of the fourth-generation prototype #3 SiGe HBT

Parameters	300 K	112 K	4.5 K
Peak $\beta$	827	6,504	7,693
Peak $g_m$ (mS)	72	110	113
Peak $f_{max}$ (GHz)	343	434	618
Peak $f_T$ (GHz)	309	403	463
Transit time $\tau_f$ (fs)	420	330	300
$V_{BE}$ at peak $f_{max}$ (V)	0.90	1.04	1.06
$I_C$ at peak $f_{max}$ (mA)	5.6	7.9	4.8
$BV_{CEO}$ (V)	1.70	1.63	1.62
$BV_{CBO}$ (V)	5.6	5.6	5.6

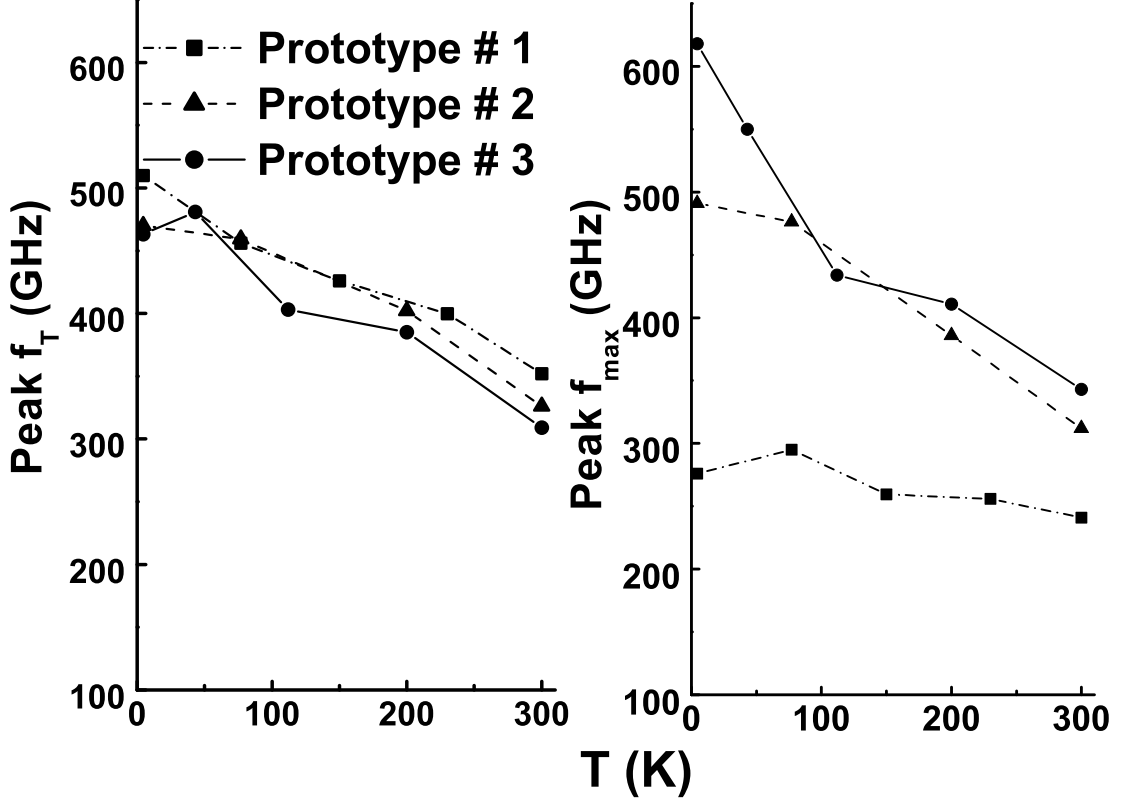
function of the chuck temperature is plotted in Fig. 52. Observe that  $\tau_{gate}$  decreases nearly monotonically with cooling. The shortest gate is 2.31 ps, achieved at 25 K, which is 0.42 ps lower than the room-temperature value (2.73 ps), an improvement of 15.4%. Decreased  $R_B$  and the natural improvement in  $f_T$  with cooling contribute to the higher speed at cryogenic temperatures, as will be discussed in the next section.

Fig. 53 shows the measured CML ring oscillator  $\tau_{gate}$  vs.  $I_{gate}$  at 25, 77, 170, and 294 K, showing the expected power-delay performance. The minimum  $\tau_{gate}$  occurs at  $I_{gate}=1.875$  mA at each temperature. Comparing the 294 K curve with the 25 K curve, one can easily see that  $I_{gate}$  drops from 1.875 mA to 1.000 mA at the same delay of 2.73 ps, effectively reducing the power consumption by one-half for the same speed. This natural reduction in power consumption with cooling represents a favorable trade for many system applications.

#### 4.3.2 Self-Heating Across Temperature

Self-heating has become increasingly important in highly-scaled devices and circuits, and we examine the self-heating-induced temperature increase in these ring oscillators.

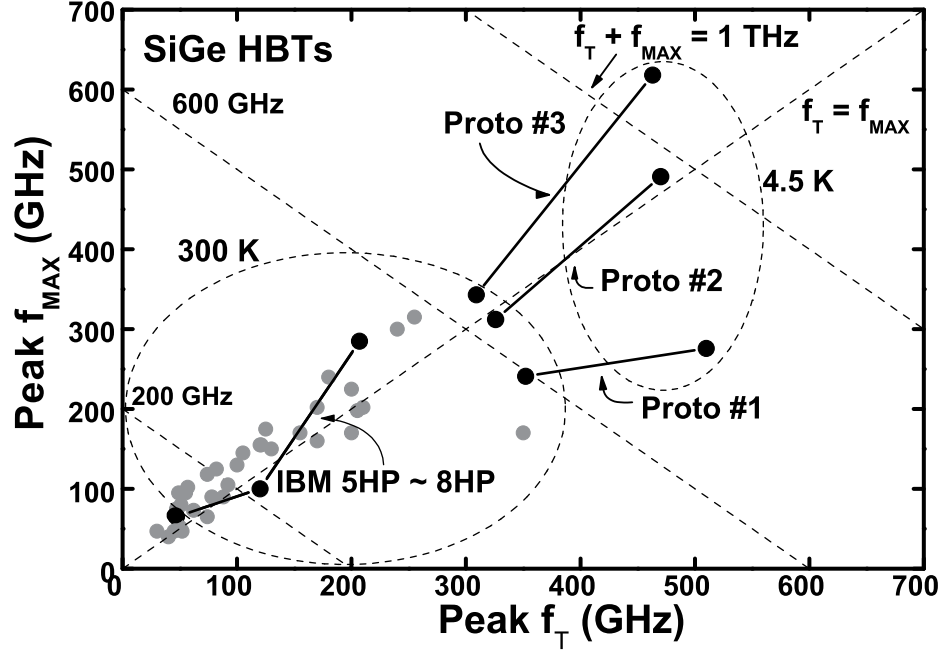
We consider self-heating at  $I_{gate}=1.875$  mA, the bias at which the shortest gate delay occurs at all temperatures. There are two components impacting any self-heating-induced temperature increase:  $\Delta T_{total} = \Delta T_1 + \Delta T_2$ , where  $\Delta T_1$  is the increase of the average temperature in the circuit area in the silicon wafer ( $T_{RO}$ ) relative to the chuck



**Figure 49:** Peak  $f_T$  and  $f_{max}$  as functions of the chuck temperature for all the three prototypes of the fourth-generation SiGe HBTs.

temperature ( $T_{chuck}$ ) as the circuit dissipates the power consumption to the backside chuck, and  $\Delta T_2$  is the increase of the SiGe HBT temperature ( $T_{HBT}$ ) relative to its local environment ( $T_{RO}$ ) as the transistor dissipates its heat to the circuit area.

The thermal conductivity of silicon ( $\kappa$ ) is dominated by point defect (isotopes) and Umklapp scattering at room temperature. However, diffuse boundary scattering is more dominant when operating below 100 K [156], and such scattering becomes increasingly significant with higher doping. For the p-type substrate in the current technology (doping is about  $10^{15} \text{ cm}^{-3}$ ), one can obtain the corresponding  $\kappa$  at 2 to 200 K from [157]. The room-temperature  $\kappa$  is known to be  $1.46 \text{ W/cm} \cdot \text{K}$ . It increases first with cooling, reaching  $15 \text{ W/cm} \cdot \text{K}$  at 50 K, and then decreases at even lower temperature. At 25 K where the fastest  $\tau_{gate}$  is achieved,  $\kappa = 13.3 \text{ W/cm} \cdot \text{K}$ , about



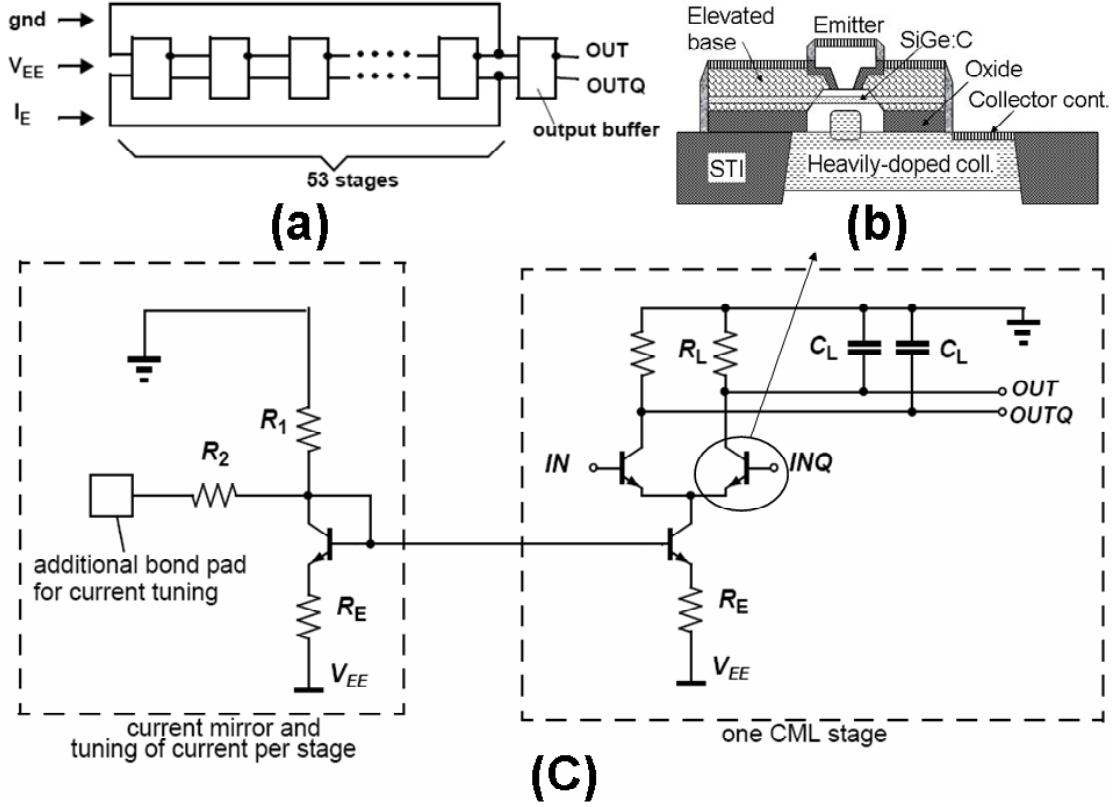
**Figure 50:** Relative location of the present result on the  $f_T$ - $f_{max}$  plane in comparison with other reported values of SiGe HBTs.

10x higher than its room-temperature value, leading to significantly lower self-heating at cryogenic temperatures, which is clearly an advantage.

Assuming  $T_{RO}$  is known, one can estimate  $\Delta T_1$  based on the layout information. The shape of the circuit corresponds to the edge of a square and with a line width ( $W$ ) of  $25 \mu m$ . Therefore it can be regarded as a combination of four rectangles with  $L = 115 \mu m$  and  $W = 25 \mu m$ . Approximating heat flow as half cylindrical at radii  $< L/2$  for each rectangle and as hemispherical at greater distances for the entire circuit, the  $\Delta T_1(T_{RO})$  relation can then be:

$$\Delta T_1 = \frac{I_{EE} V_{EE}}{\pi \kappa (T_{RO}) L} \left( \frac{1}{4} \ln \left( \frac{L}{W} \right) + 1 \right), \quad (22)$$

where  $I_{EE}$  and  $V_{EE}$  are the bias current and voltage. Meanwhile,  $T_{RO}$  can also be



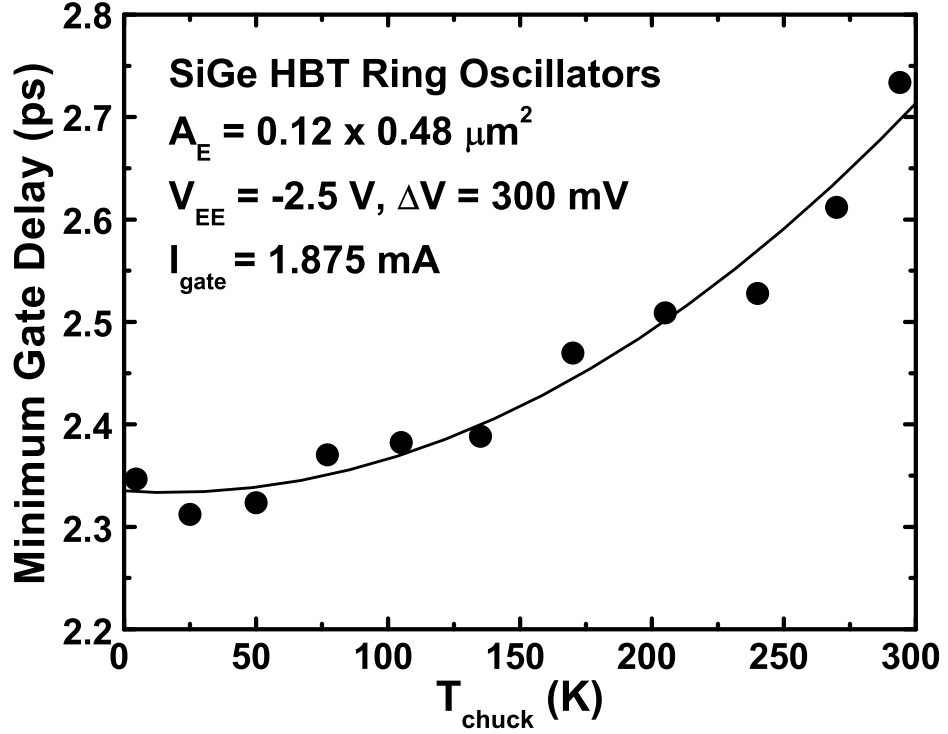
**Figure 51:** Schematic of (a) the CML ring oscillator, (b) cross-section of the SiGe HBT structure, and (c) current mirror and one CML gate of the ring oscillator.

expressed as:

$$T_{RO} = T_{chuck} + \Delta T_1. \quad (23)$$

Numerical iteration based on Eqs. (22) and (23) and the  $\kappa(T_{RO})$  yields a precise calculation of  $\Delta T_1$  as a function of  $T_{chuck}$  (Fig. 54). Self-heating increases the average temperature in the circuit area by 7.6 K when  $T_{chuck} = 294$  K; however,  $\Delta T_1$  is only 0.73 K when  $T_{chuck} = 25$  K.

$\Delta T_2$  is determined by the n-type sub-collector that is doped around  $10^{19} \text{ cm}^{-3}$  to  $10^{20} \text{ cm}^{-3}$ . The temperature dependence of such heavily-doped n-type silicon can be obtained from [158]. The low-temperature  $\kappa$  is now dramatically decreased

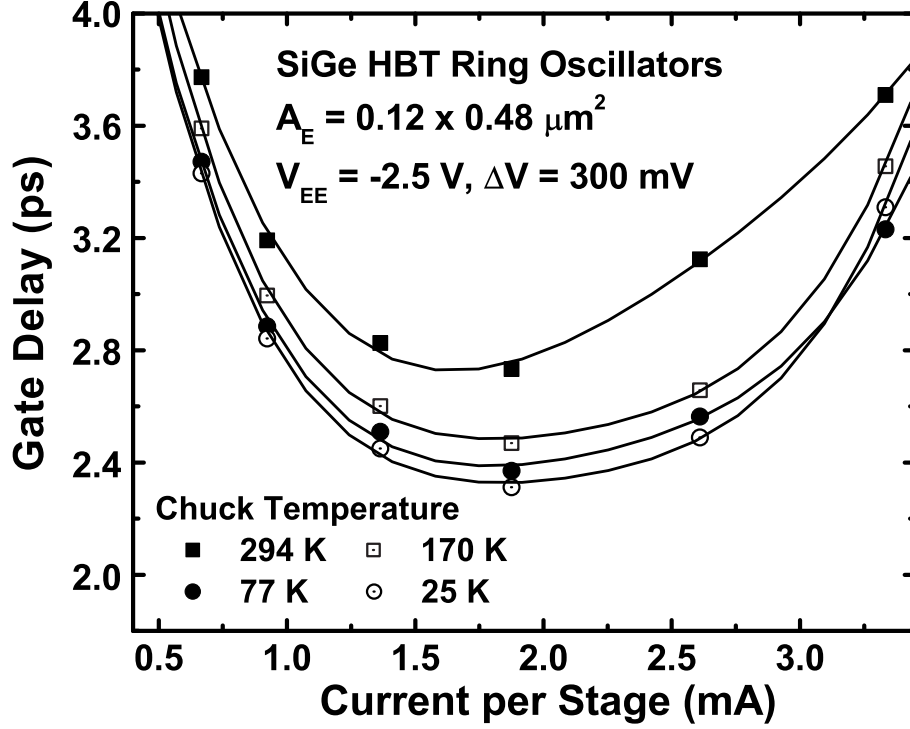


**Figure 52:** Measured minimum CML ring oscillator gate delay as a function of chuck temperature.

compared to lightly-doped Si because of stronger boundary scattering. Given that the transistor thermal resistance ( $R_{th}$ ) is 11.5 K/mW at room temperature, one can scale  $R_{th}$  inversely proportional to  $\kappa$  across the entire temperature range. The  $\Delta T_2(T_{HBT})$  relation can then be calculated in a similar manner as  $\Delta T_1(T_{RO})$  by assuming an average  $V_{CE} = 1 \text{ V}$ , from which the  $\Delta T_2(T_{RO})$  relation is derived. Since  $\Delta T_1(T_{chuck})$  is already known,  $\Delta T_2$  as a function of  $T_{chuck}$  can finally be calculated, and is plotted in Fig. 54.

Adding  $\Delta T_1$  and  $\Delta T_2$  leads to the final  $\Delta T_{total}(T_{chuck})$  curve plotted in the same figure. The ratio of  $\Delta T_1$  to  $\Delta T_2$  decreases with cooling, and  $\Delta T_2$  dominates  $\Delta T_{total}$  at  $T_{chuck} < 110 \text{ K}$ , because: (a)  $\kappa$  of the p-substrate increases dramatically at such low



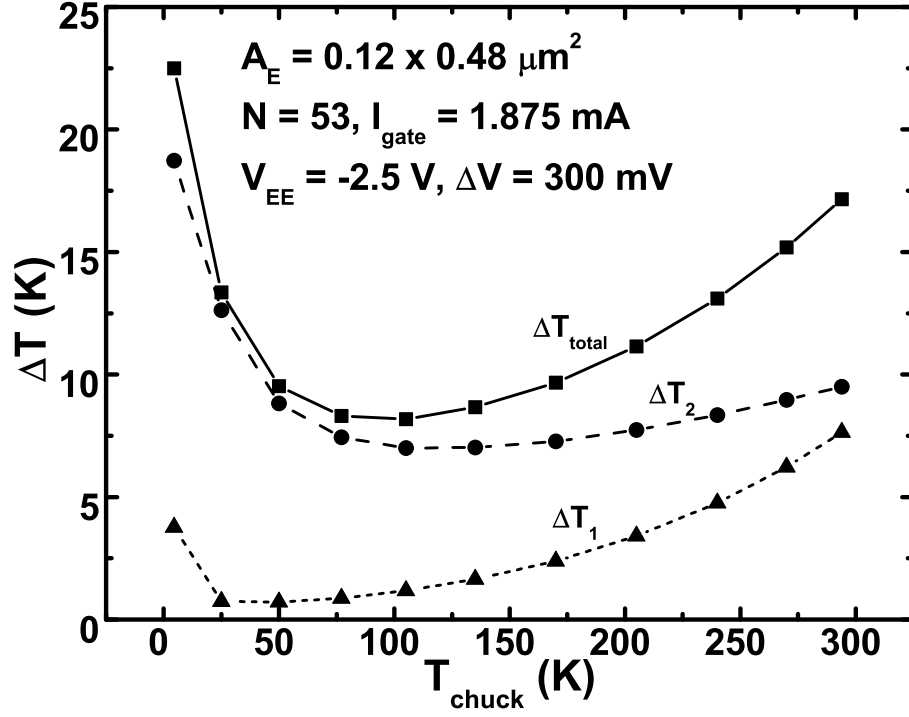


**Figure 53:** Measured CML ring oscillator gate delay *vs.* current per gate at 25, 77, 170, 294 K.

temperatures (lower  $\Delta T_1$ ), and (b) enhanced diffuse boundary scattering decreases  $\kappa$  of the highly-doped sub-collector significantly (higher  $\Delta T_2$ ). The overall  $\Delta T_{total}$  is still lower at most cryogenic temperatures, decreasing from 17.1 K when  $T_{chuck} = 294$  K, to 8.2 K when  $T_{chuck} = 105$  K, and 13.3 K when  $T_{chuck} = 25$  K, demonstrating another advantage of cryogenic operation over room temperature: less self-heating due to the higher thermal conductivity of the substrate.

#### 4.4 Temperature Dependence of Intrinsic Base Resistance with Scaling

The intrinsic base sheet resistance ( $R_{bi}$ ) was studied across temperature in Chapter 3. Severe carrier freezeout increases the low-temperature  $R_{bi}$  significantly, and

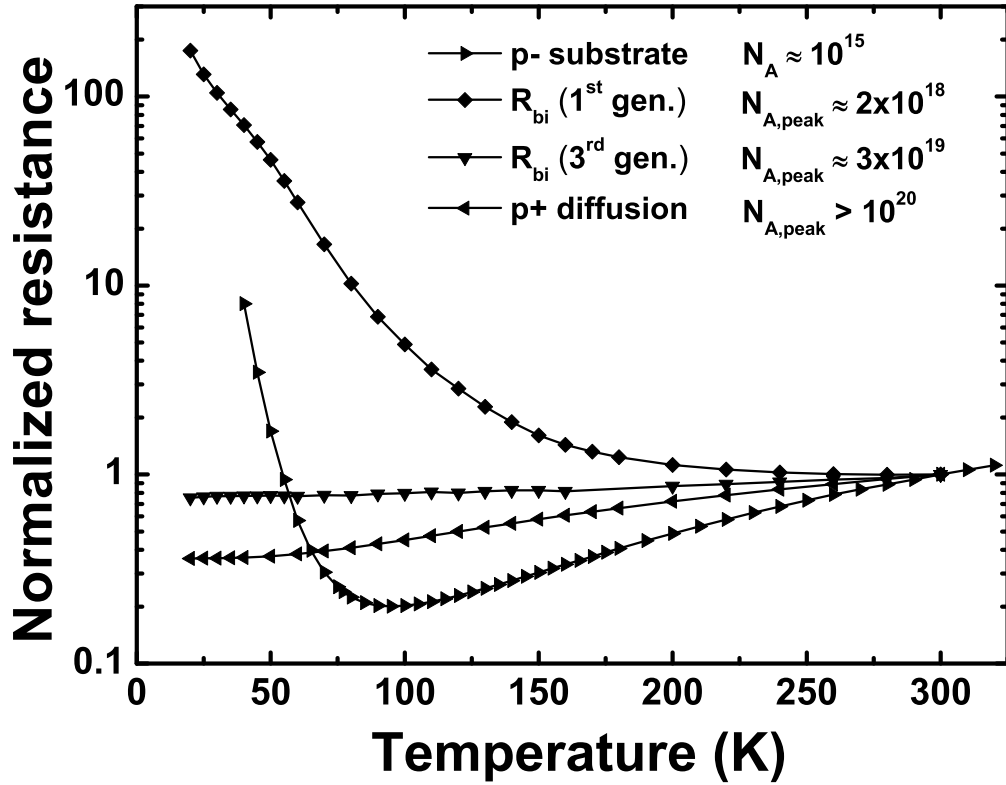


**Figure 54:** Self-heating-induced temperature increase:  $\Delta T_{\text{total}} = \Delta T_1 + \Delta T_2$ , where  $\Delta T_1$  is the increase of the average temperature in the circuit area ( $T_{RO}$ ) relative to the chuck temperature ( $T_{\text{chuck}}$ ), and  $\Delta T_2$  is the increase of the HBT temperature ( $T_{HBT}$ ) relative to  $T_{RO}$ .

this effect is expected to disappear with technology scaling. For third-generation SiGe HBTs and beyond, intrinsic base doping is higher than a few  $10^{19} \text{ cm}^{-3}$ . In this section, the temperature-dependence of  $R_{bi}$  was measured in third-generation SiGe HBTs for the first time, using the tetrode structure and the measurement technique described in [159].

In Fig. 55, the measured resistance of p-type silicon *vs.* temperature is shown for a range of doping concentrations, normalized to 300 K. As discussed in Chapter 3, for doping concentrations below the Mott-transition (e.g., in the p-substrate and base region of a first-generation NPN SiGe HBT), the temperature dependence is driven

by both the carrier mobility and dopant ionization level. For a third-generation SiGe HBT and beyond, the peak base doping (about  $3 \times 10^{19} \text{ cm}^{-3}$ ) lies above the Mott-transition; thus, incomplete ionization is negligible across temperature, and the change in base resistance across temperature is driven by the carrier mobility, which decreases by about 25% from 300 K to 20 K, giving rise to a higher speed at 25 K than at 294 K, as measured. Similarly, the highly doped diffusion layer shows a moderate decrease in mobility as the temperature decreases.



**Figure 55:** Measured temperature dependence of the normalized base sheet resistance of a third-generation SiGe HBT (peak doping about  $3 \times 10^{19} \text{ cm}^{-3}$ ) in comparison with the normalized sheet resistance of other p-type Si/SiGe layers that were shown in Fig. 24.

To understand the temperature dependence of the carrier mobility that determines the third-generation  $R_{bi}$ , one needs to look at the fundamental scattering

mechanisms, including lattice, donor, acceptor, and electron-hole scattering [160], [161]. For increasing impurity concentration, impurity and carrier scattering dominate the mobility in the temperature range of interest [147]. The expression for the combined impurity and carrier scattering mobility component consists of two terms, the first of which has a power law dependence on temperature and dominates below the Mott-transition, where carrier screening of impurities is weak. Above the Mott-transition, carrier screening becomes important and the carrier mobility is dominated by the second term, which has an inverse power law dependence on temperature in the absence of carrier freezeout. Since in this case the dopants remain totally ionized across temperature, with increasing doping the total carrier mobility approaches an inverse power law dependence and exhibits a moderate increase as the temperature decreases. Consequently, with complete dopant ionization,  $R_{bi}$  merely reflects the inverse of the change in mobility with cooling, consistent with our data.

## 4.5 *Summary*

We have investigated the path towards achieving terahertz speeds in SiGe HBTs by examining how device scaling techniques couple the cryogenic DC and AC characteristics of these transistors. Cooling has been demonstrated to be qualitatively similar to vertical and lateral device scaling with respect to DC, AC, and noise performance in SiGe HBTs. Three different prototype fourth-generation devices were used as a case study to explore the performance limits associated with vertical scaling and parasitic reduction. A record  $f_{max}$  for a silicon-based transistor and the first combined set of  $f_T$  and  $f_{max}$  above one-terahertz for SiGe HBTs were achieved. Peak  $f_{max}$  of 618 GHz and  $f_T$  of 463 GHz at 4.5 K (chuck temperature) were measured for a SiGe HBT (343 / 309 GHz at 300 K), at  $BV_{CEO}$  of 1.62 V at 4.5 K (1.70 at 300 K) and  $BV_{CBO}$  of 5.6 V at both 300 and 4.5 K, yielding a record  $f_T \times BV_{CEO}$  product of 750 GHz-V (510 GHz-V at 300 K). We have also demonstrated a record

2.3 ps CML gate delay at 25 K in SiGe ring oscillators operating at 25 K. The lower  $R_{bi}$  and the higher  $\kappa$  at cryogenic temperatures make SiGe a compelling technology for low-temperature ultra-high-speed digital applications.

Cooled SiGe HBTs are intriguing for a number of high-performance cryogenic applications mentioned in Chapter 1. For modern space systems (e.g., as needed for Earth orbit, upcoming lunar and Martian missions, as well to the outer planets such as Jupiter and Saturn and their moons), the extremely cold ambient environment encountered (e.g., -230 C or 43 K in the polar craters on the surface of the Moon) is highly detrimental to most electronic systems. As demonstrated here once again, operation in such extreme cold environments is a natural fit for SiGe devices since their performance intrinsically improves with cooling. All of these results bode well for the further scaling of SiGe HBTs towards room temperature THz speeds. The research work on the cryogenic SiGe HBTs helps to identify the key components for future scaling for THz performance at 300 K.

## CHAPTER V

### SCALING FOR TERAHERTZ SPEEDS

Although an effective tool for enhancing speed in SiGe HBTs as evidenced in the previous chapter, cooling comes at added complexity and cost, and its application is restricted due to the availability of refrigeration systems. In this chapter, we investigate two other approaches to reach terahertz speeds: alternative circuit topologies and device scaling.

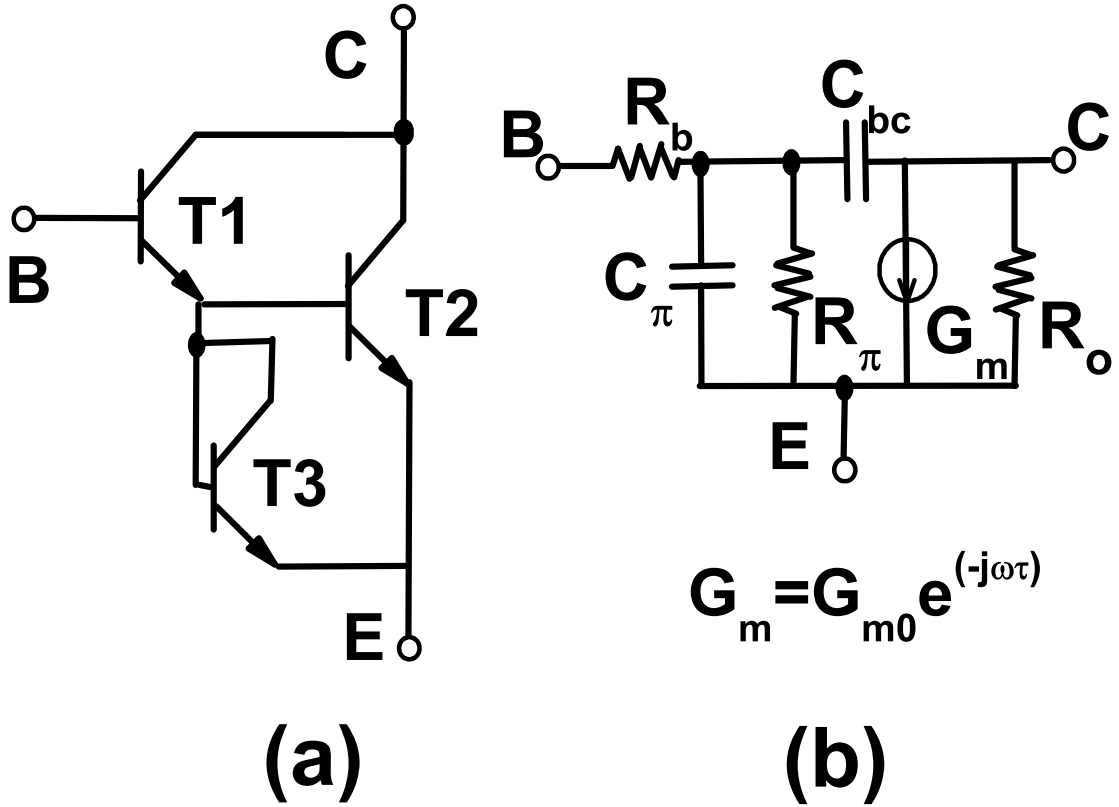
#### ***5.1 The $f_T$ -Doubler Technique***

As a low-cost alternative, we here explore the use of novel circuit techniques to enhance speed, without changing the underlying core process technology. As a prototype example, it is well known that the bandwidth of a circuit can be increased by utilizing  $f_T$ -doubler techniques [162]-[164]. The potential use of  $f_T$ -doubler topologies in SiGe technology have been explored in first-generation SiGe HBTs [165], and second-generation SiGe HBTs for application to high-speed A/D converters [166].

In the present section, a 325 GHz  $f_T$ -doubler circuit is implemented, for the first time, in the third-generation, 130 nm, 200 GHz SiGe HBT. Reduced  $C_\pi$  is shown to be mainly responsible for the improvement in the cutoff frequency. Small-signal equivalent circuit is investigated, and we show that the  $f_T$ -doubler can be treated as a single transistor unit cell during circuit design, making it easy to incorporate into actual circuits. The impact of emitter geometry and temperature on the performance are investigated. This SiGe  $f_T$ -doubler reaches a record peak  $f_T$  of 438 GHz at 93 K.

### 5.1.1 $f_T$ -Doubler Analysis

The core of the SiGe HBT  $f_T$ -doubler is identical to a classical Darlington pair, as shown in Fig. 56(a). The series connection of T1 and T2 reduces the input capacitance by one-half for each transistor. T3 biases T1 and T2 roughly at the same current level to make the  $f_T$ 's of the devices approximately equal. The presence of T3 in parallel with T2, however, increases the input capacitance, slightly degrading  $f_T$ . Therefore the  $f_T$ -doubler cannot, in practice, truly double the  $f_T$  of the core transistor building block.



**Figure 56:** (a) Schematic of the  $f_T$ -doubler; (b) small-signal equivalent circuit of a SiGe HBT.

Using the small-signal equivalent circuit of a single SiGe HBT (Fig. 56(b)), and assuming an ideal situation where  $R_b = 0$  and  $1/R_\pi = 0$ , one obtains the current gain,  $h_{21}$  of the doubler as,

$$h_{21} = \frac{2g_m}{j\omega(C_\pi + 2C_{bc})}. \quad (24)$$

Note that  $f_T$  of a single transistor can be expressed as,

$$f_{T,HBT} = \frac{g_m}{2\pi(C_\pi + C_{bc})}. \quad (25)$$

Hence, a more accurate expression for an  $f_T$ -doubler is,

$$f_{T,doubler} = \left(1 + \frac{C_\pi}{C_\pi + 2C_{bc}}\right) f_{T,HBT}. \quad (26)$$

Note that the above derivation is established upon the assumption that  $C_\pi$  and  $C_{bc}$  are the same for all three transistors, which is not strictly true since T3 has  $V_{CB} = 0$  V. Equation (25), nevertheless, gives fundamental insight into the operating principles of  $f_T$ -doublers. In addition, the base resistance of the  $f_T$ -doubler is approximately  $2R_{b,HBT}$ , negating the positive influence of the reduction of input capacitance on  $f_{max}$ .

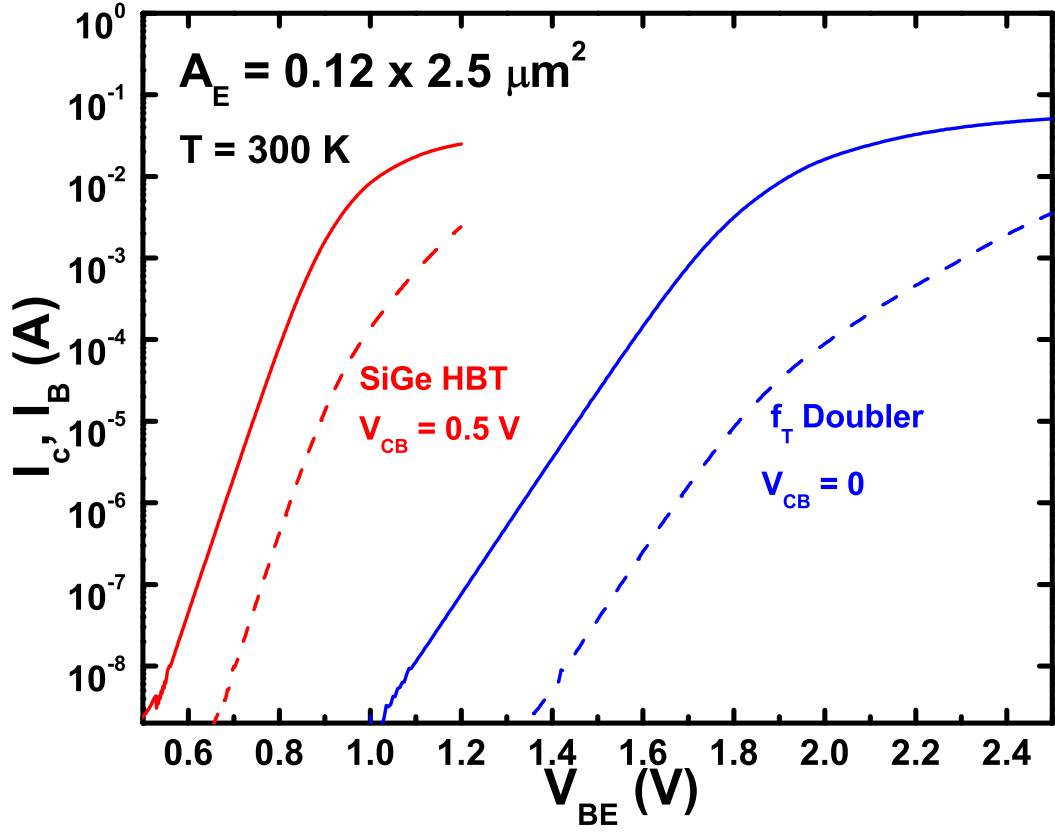
### 5.1.2 DC and AC Characteristics

All devices and circuits were fabricated using IBM 8HP SiGe BiCMOS technology. The DC measurements were performed for both a single  $0.12 \times 2.5 \mu m^2$  SiGe HBT and a SiGe HBT  $f_T$ -doubler with the same underlying emitter geometry. The forward Gummel characteristics are shown in Fig. 57, from which  $\beta$  and  $g_m$  can be extracted. The  $f_T$ -doubler produces more than 2x the  $\beta$  at the same  $I_C$  level, while maintaining similar  $g_m$  values (Fig. 58).

Output characteristics of the single transistor and the  $f_T$ -doubler are presented in Fig. 59, with  $V_{CE} = 1.5$  V and 2.25 V, respectively. The collector voltage shift of the  $f_T$ -doubler, due to T3, is about 0.75 V. Observe that  $I_C$  is almost doubled at a comparable  $I_B$  input drive.

Standard LRM calibration and de-embedding methods were used and the S-parameters measured for the SiGe HBT and the  $f_T$ -doubler, up to 40 GHz. After  $h_{21}$

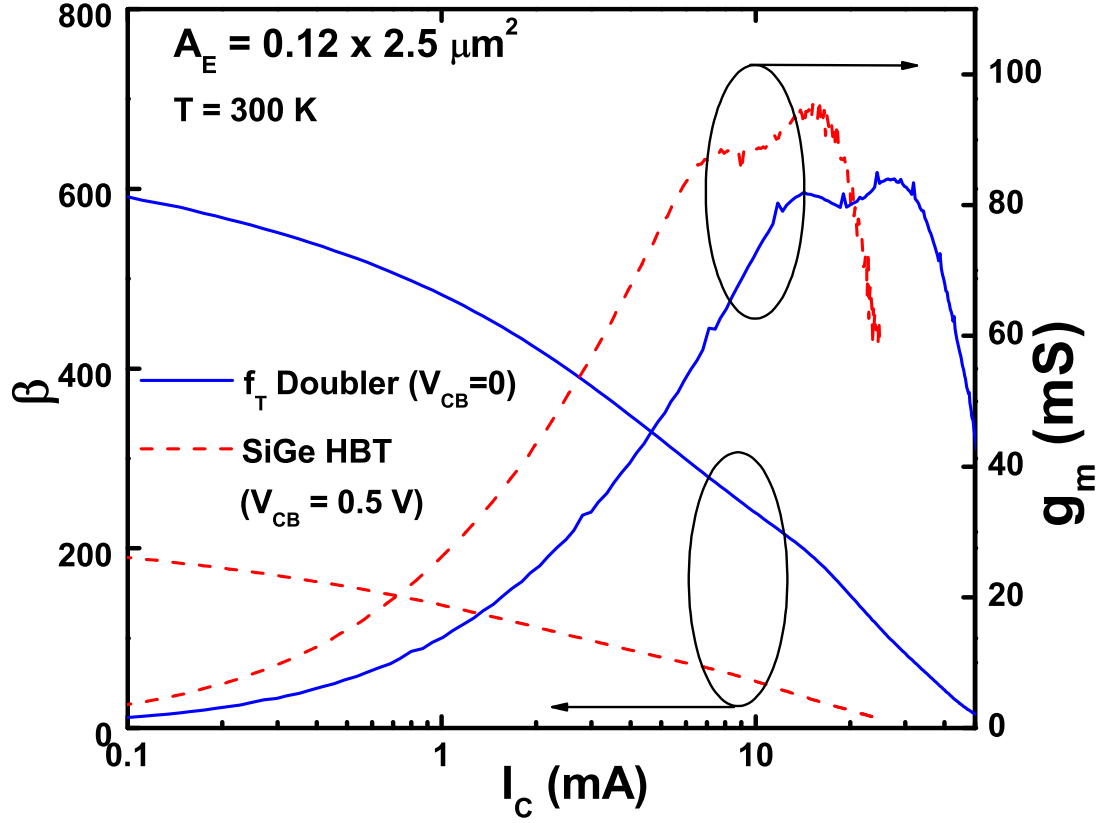




**Figure 57:** Measured forward Gummel characteristics of a third-generation SiGe HBT and a SiGe HBT  $f_T$ -doubler, using an emitter geometry of  $0.12 \times 2.5 \mu\text{m}^2$

and maximum available gain ( $MAG$ ) were calculated,  $f_T$  and  $f_{max}$  can then be extrapolated from 40 GHz, as depicted in Fig. 60. The peak  $f_T$  increases from 196 GHz (single SiGe HBT) to 325 GHz ( $f_T$ -doubler), while  $f_{max}$  and  $MAG$  show moderate decreases.

Small-signal current gain ( $h_{21}$ ) and  $MAG$  at peak- $f_T$  bias are plotted in Fig. 61 (the  $U$  data was not stable, consistent with [165]). The -20 dB/dec slope of  $h_{21}$  clearly indicates the validity of our  $f_T$  extraction. However, our measured frequency range is too low to reach the -20 dB/dec region for the  $MAG$  data, indicating that the extracted  $f_{max}$  in Fig. 60 is overly conservative.

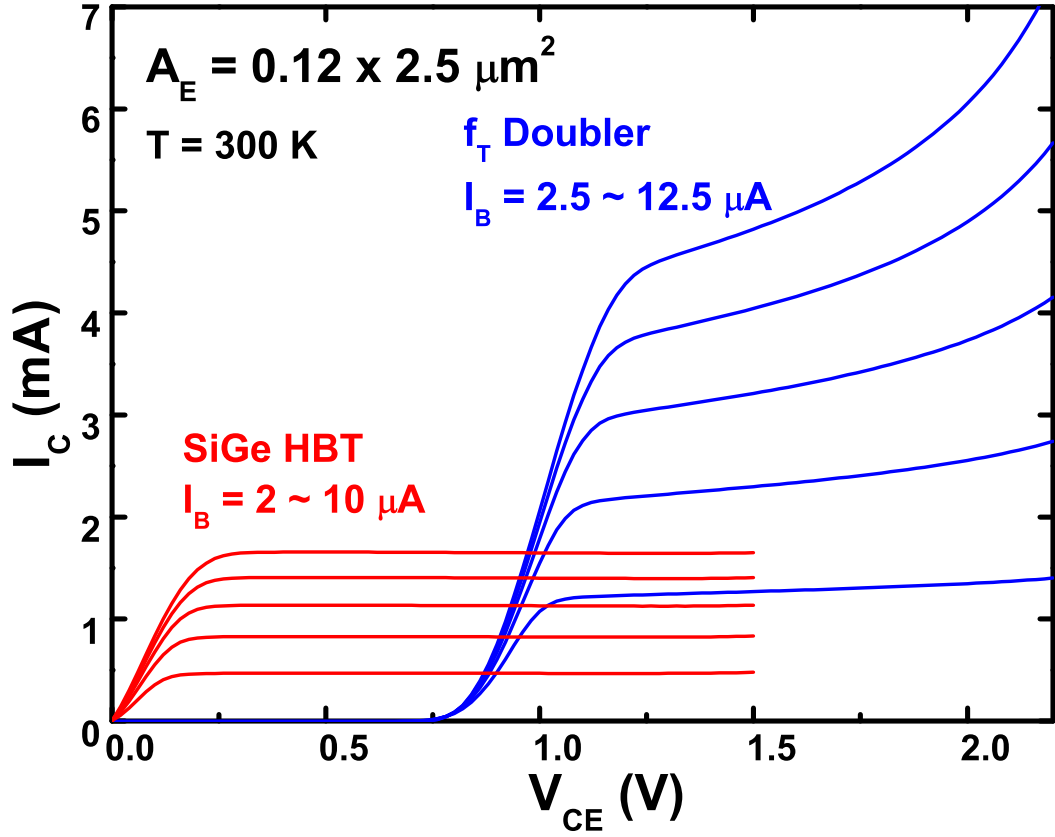


**Figure 58:** DC current gains ( $\beta$ ) and transconductances ( $g_m$ ) vs.  $I_C$  extracted from DC characteristics of the SiGe HBT and the  $f_T$ -doubler.

### 5.1.3 Small-Signal Equivalent Circuit

To better understand operation of the  $f_T$ -doubler, the important small-signal parameters in Fig. 56(b), including  $R_b$ ,  $R_\pi$ ,  $R_O$ ,  $C_\pi$ ,  $C_O$ , and  $G_{m0}$ , were extracted for the  $0.12 \times 2.5 \mu m^2$  SiGe HBT and the  $f_T$ -doubler built from the same device (Figs. 62-64). The validity of this extraction method in SiGe HBTs has been well-established in [167], and this method is also assumed to hold for the  $f_T$ -doubler in the current work. This assumption can be verified by comparing the measured and modeled S-parameters (Fig. 65), which show excellent agreement.

Comparing these parameters across all bias conditions, and especially at peak- $f_T$  biases shown in Table 4, one is easily to find the advantages of using the  $f_T$ -doubler

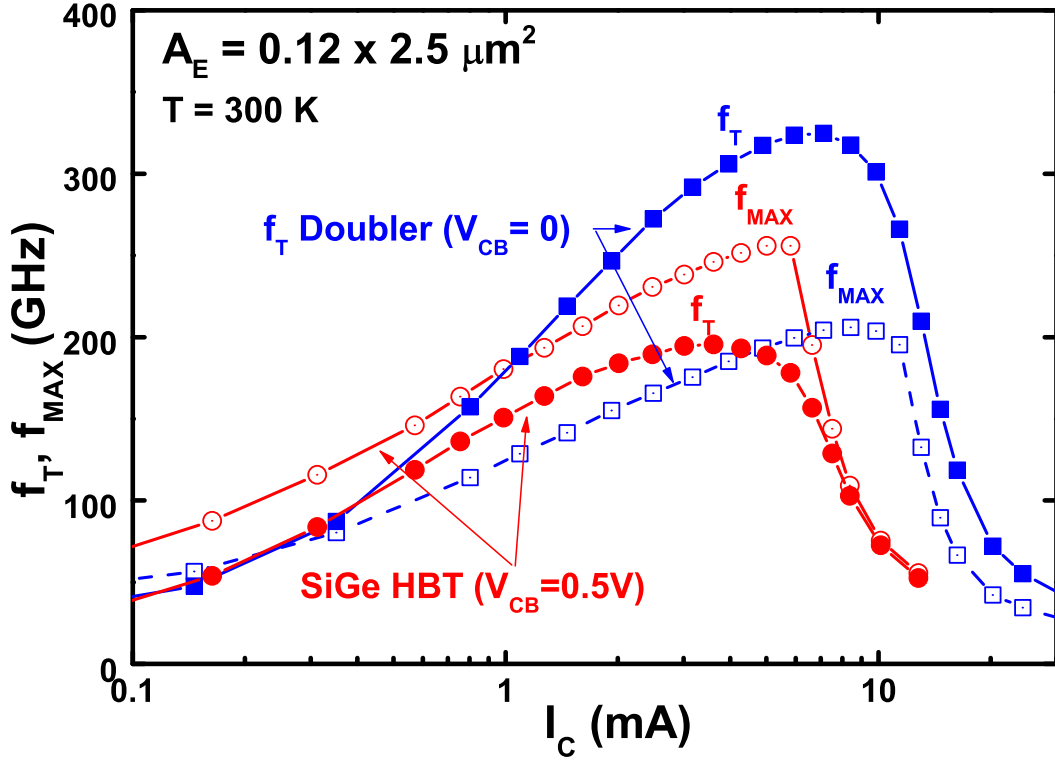


**Figure 59:** Forced- $I_B$  output characteristics of the SiGe HBT and the  $f_T$ -doubler.

topology. Small-signal transconductance ( $G_{m0}$ ) of the  $f_T$ -doubler is almost the same as that of a single transistor at peak- $f_T$  biases (Fig. 62). Figures 63 and 64 indicate that  $C_{bc}$  increases a little while  $C_\pi$  is greatly reduced for the  $f_T$ -doubler, resulting in a decrease in the total input capacitance  $C_\pi + C_{bc}$  (from 35.8 fF down to 20.8 fF), and therefore  $f_T$  is increased.  $R_\pi$ , is also improved (Figures 63).

However, side effects exist in this topology:  $R_b$  almost doubles, as expected, negating the positive influence of the  $f_T$  increase on the total power gain and thus on  $f_{max}$ . In addition,  $R_O$  is nearly halved.

Despite the modest degradation in the power gain, and hence  $f_{max}$ , the greatly enhanced current gain and  $f_T$  in the  $f_T$ -doubler topology is definitely useful for circuit

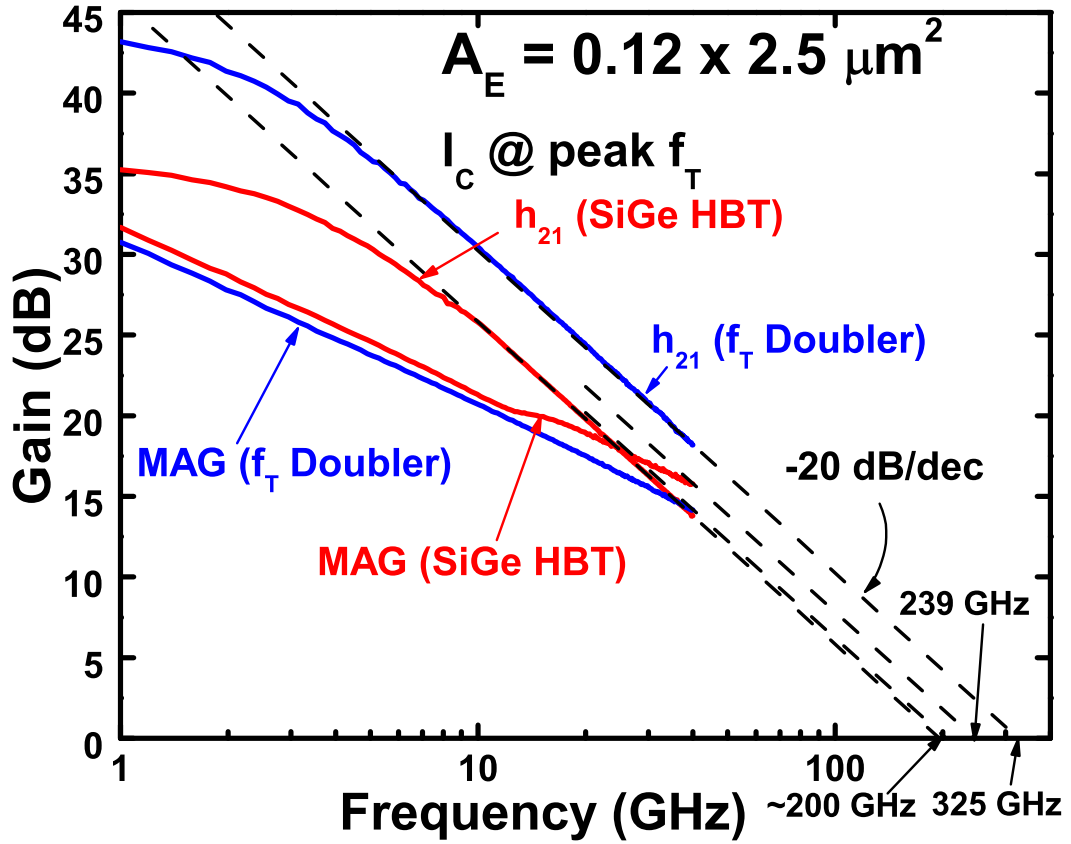


**Figure 60:**  $f_T$  and  $f_{max}$  as functions of  $I_C$  of the SiGe HBT and the  $f_T$ -doubler extrapolated at 40 GHz.

designers in a wide class of circuit applications. Circuits such as VCOs, ADC cores, frequency divider/multipliers, and circuit output stages, where current gain is more important than power gain, can all potentially benefit from this doubler topology. In addition, the application of the  $f_T$ -doubler techniques to the inherently slower and lower gain PNP SiGe HBTs in modern complementary BiCMOS technologies can provide better matching with NPN SiGe HBTs, producing more-ideal push-pull circuit configurations.

#### 5.1.4 Emitter Length Dependence

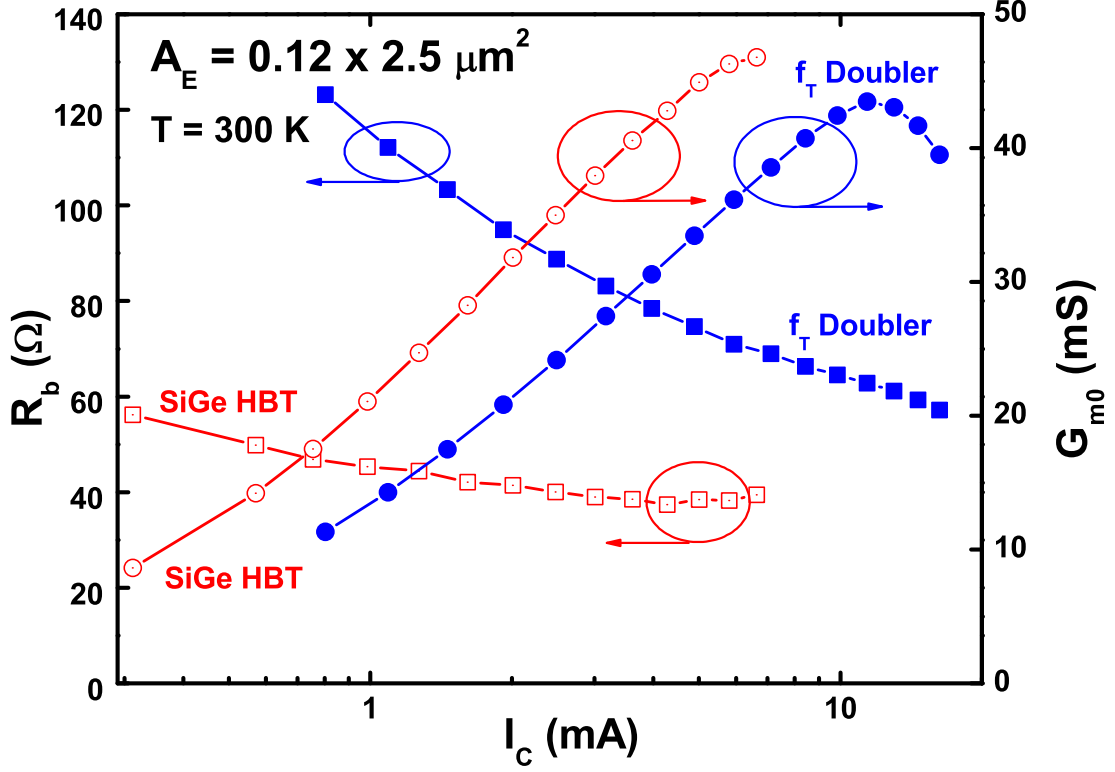
Emitter geometry (width  $W_E$  and length  $L_E$ ) is known to have an impact on the frequency response of SiGe HBTs, and the best performance is reported on the  $0.12 \times 2.5 \mu m^2$  SiGe HBTs in the IBM 8HP SiGe technology. While the impact of



**Figure 61:** Maximum available gain ( $MAG$ ) and  $h_{21}$  of the  $f_T$ -doubler and the SiGe HBT biased at a collector current needed to attain peak  $f_T$ .

$W_E$  is straightforward, since larger  $W_E$  increases the total  $R_b$  and hence degrades  $f_{max}$ , the role of  $L_E$  is more complicated, as it is affected by process issues and other effects such as fringing fields and current crowding, and thus is worth re-visiting in this context.

The peak  $f_T$  as a function of  $W_E$  for single SiGe HBTs and  $f_T$ -doublers are measured and compared in Fig. 66, in which  $f_T$ -doublers consistently show more than a 50% speed improvement and the  $0.12 \times 2.5 \mu\text{m}^2$  emitter size remains the best choice for this SiGe technology.



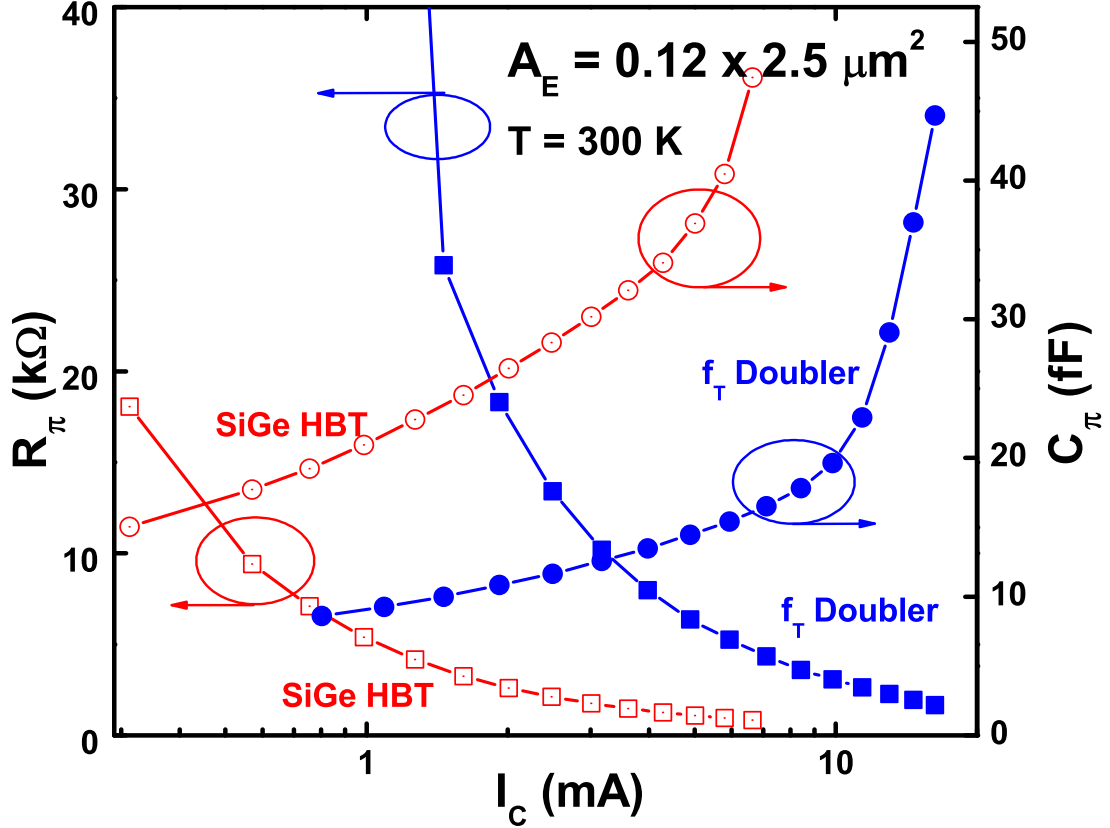
**Figure 62:** Extracted  $R_b$  and  $G_{m0}$  as functions of  $I_C$  for the SiGe HBT and the  $f_T$ -doubler.

### 5.1.5 Temperature Effects

Cryogenic temperatures help enhance the speed of the  $f_T$ -doubler, just as it does for the single device (Fig. 67). A record  $f_T$  of 438 GHz is achieved at 93 K for the doubler, and the small-signal parameters at 93 K have also been derived (Table 4). Obviously, the increased  $G_{m0}$  and the decreased total input capacitance both act to enhance the speed at cryogenic temperatures.

## 5.2 New Scaling Rule and Roadmap

Conventional (and simple) scaling of Si bipolar transistors ended in the 1990s when the industry moved to CMOS devices, mainly because of the challenges associated with power density limits in conventional bipolar circuits. It was predicted in the literature that one would increase the collector current density ( $J_C$ ) by approximately

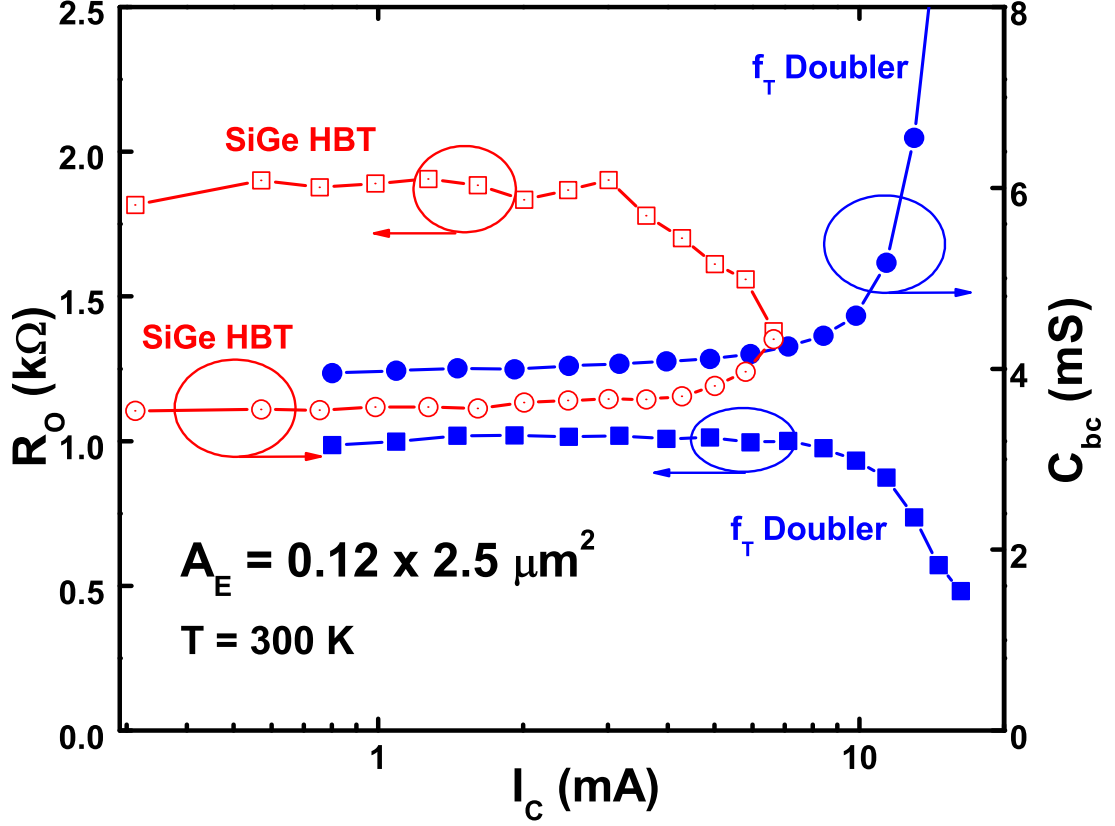


**Figure 63:** Extracted  $R_\pi$  and  $C_\pi$  as functions of  $I_C$  for the SiGe HBT and the  $f_T$ -doubler.

$\lambda^2$  in order to improve the device speed by  $\lambda$  [168]-[171]. Even today's InP HBTs are still following this scaling rule [90]. The incorporation of bandgap engineering with SiGe revived the scaling of Si bipolar transistors. However, as the  $J_C$  at peak  $f_T$  already hits 19 mA/ $\mu\text{m}^2$  in the fourth-generation SiGe HBTs, such  $J_C \propto \lambda^2$  scaling rule can soon become problematic for reliability.

Based on the work [63] on the statistics of SiGe HBTs produced in the past decade, we propose a new scaling rule and a new roadmap for future SiGe HBTs, as summarized in Table 5. At 32 nm lithography node, SiGe HBTs should in principle achieve THz speeds with  $f_T / f_{max}$  of 782 / 910 GHz and  $BV_{CEO}$  of 1.1 V.

Compared to conventional scaling rules mentioned in  $\lambda$  [168]-[171],  $J_C$  at peak  $f_T$  here is proportional to  $\lambda^{1.35}$ , showing the reluctance of SiGe designers to increase

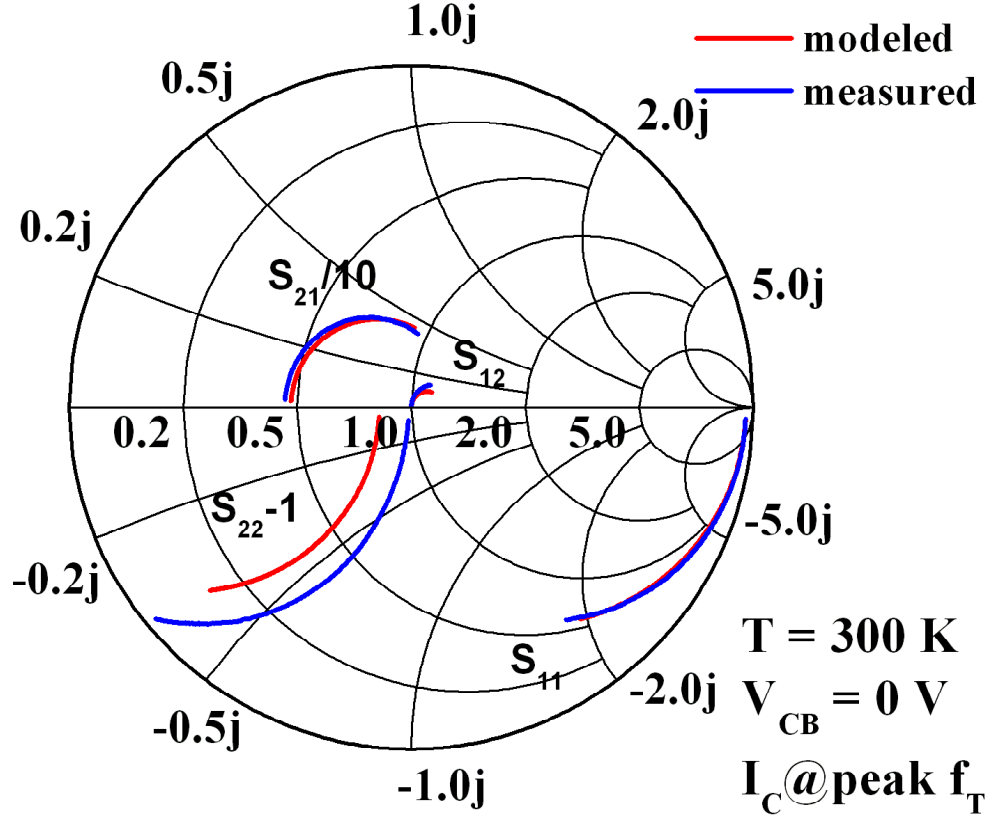


**Figure 64:** Extracted  $R_O$  and  $C_{bc}$  as functions of  $I_C$  for the SiGe HBT and the  $f_T$ -doubler.

collector doping too fast, for several reasons: self-heating, electro-migration, and breakdown voltages. Up to now, the speed of SiGe HBTs has been mainly limited by Kirk effect (base push-out) and heterojunction barrier effect (HBE) [19]. Careful device engineering employed in state-of-the-art SiGe HBTs improves the slope of the  $f_T \times BV_{CEO}$  trend in SiGe HBTs, in contrast to that for III-V HBTs (Fig. 3), and a record  $f_T \times BV_{CEO}$  of 750 GHz-V is demonstrated in Si-based transistors, benefiting from the powerful tool of cooling in enhancing the frequency response of SiGe HBTs, with only a minimal degradation in breakdown voltage.

However, as the collector doping in SiGe HBTs is further increased to suppress Kirk effect and HBE for better frequency response, the impact-ionization rate at the collector-base junction increases, and an inherent (and well-known) tradeoff exists





**Figure 65:** Comparison of the measured and modeled S-parameters ( $S_{11}$ ,  $S_{12}$ ,  $S_{21}/10$ , and  $S_{22}$ ) for the  $f_T$ -doubler.

between peak  $f_T$  and breakdown voltage [172],[119]. The  $f_T \times BV_{CEO}$  will eventually be limited by the physical upper bound of  $E_{max}v_{sat}/\pi$  where  $E_{max}$  is the maximum electrical field that the BC junction can sustain. Theoretically, this limit can be slightly exceeded because  $W_{CSCL}$  is an increasing function of  $V_{CB}$  bias. Therefore, the  $V_{CB}$  where peak  $f_T$  is achieved in the highly-scaled SiGe HBTs will eventually be lower than the  $V_{CB}$  when  $BV_{CEO}$  is measured. Consequently, clever circuit designers can optimize the circuit topology to achieve the best needed  $f_T$  and  $BV_{CEO}$ , respectively, by adjusting the corresponding  $V_{CB}$ . In spite of this, the breakdown voltage is becoming one of the major bottlenecks for SiGe HBT scaling. Engineering the intrinsic layer of the BC junction, the doping species of the SIC region [87], and the Ge retrograde in the BC junction, therefore, becomes critically important. Meanwhile,

**Table 4:** Small-signal parameters for a  $0.12 \times 2.5 \mu\text{m}^2$  third-generation SiGe HBT and an  $f_T$ -doubler biased at peak  $f_T$

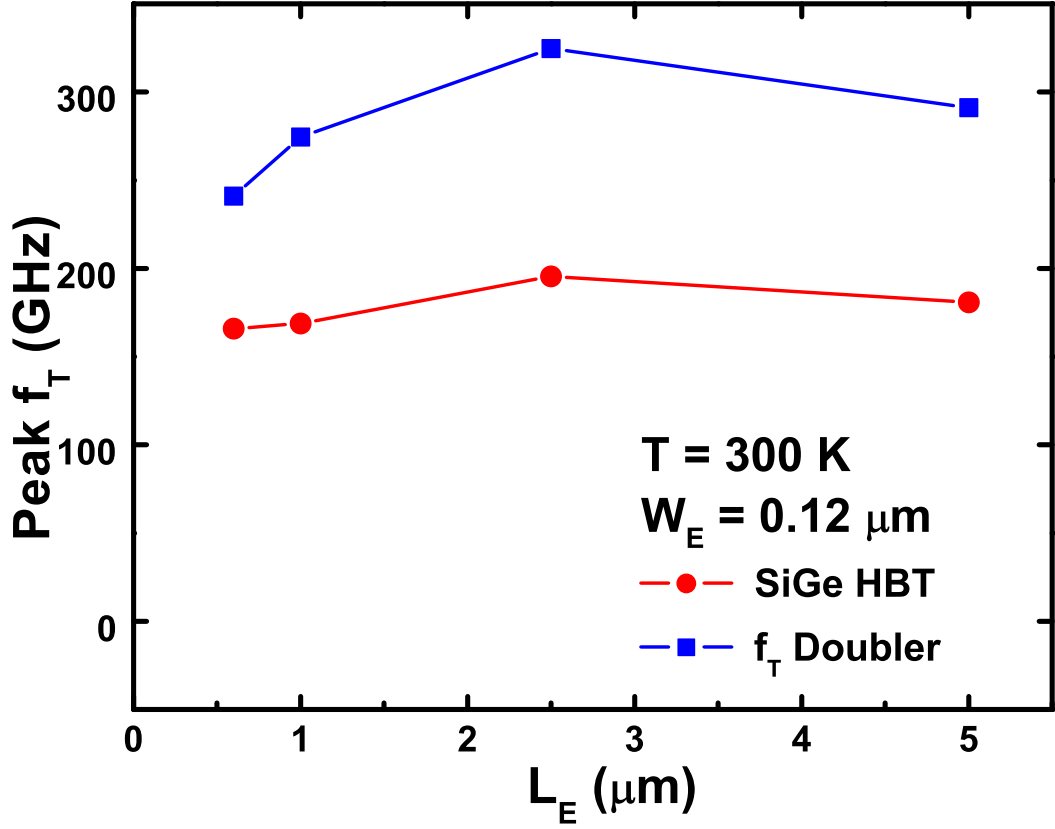
	SiGe HBT (300 K)	$f_T$ -Doubler (300 K)	$f_T$ -Doubler (93 K)
$f_T$ (GHz)	196	325	438
$I_C$ (mA)	3.61	7.12	6.45
$R_b$ ( $\Omega$ )	38.5	69.0	62.0
$R_\pi$ ( $\Omega$ )	1474.2	4348.1	2488.9
$R_O$ ( $\Omega$ )	1778.8	1000.8	1327.4
$C_{bc}$ (fF)	3.66	4.25	4.00
$C_\pi$ (fF)	32.09	16.52	17.31
$G_{m0}$ (fF)	40.6	38.6	59.1

**Table 5:** New SiGe HBT scaling rule and roadmap

Emitter	Emitter Width ( $nm$ )	$\lambda^{-1.4}$	120	90	65	45	32
Base	$R_B$ ( $\Omega$ )	$\lambda^{-0.05}$	34.5	34	33.5	33	32.5
Collector	$C_{CB}$ (fF)	$\lambda^{-0.95}$	4.93	4.06	3.26	2.54	2.02
	Collector Width ( $\mu\text{m}$ )	$\lambda^{-1.4}$	0.40	0.32	0.26	0.20	0.15
	$J_C$ at peak $f_T$	$\lambda^{1.35}$	19	25	34	49	68
Performance	$f_T$ (GHz)	$\lambda$	300	370	470	610	782
	$f_{max}$ (GHz)	$\lambda$	350	430	550	710	910
	$BV_{CEO}$ (V)	$\lambda^{-0.45}$	1.70	1.55	1.40	1.25	1.10
	$BV_{CBO}$ (V)	$\lambda^{-0.45}$	5.6	5.1	4.6	4.1	3.6
	$\tau_f$ (ps)	$\lambda^{-1}$	0.42	0.34	0.27	0.21	0.16
	$\tau_{gate}$ (ps)	$\lambda^{-1}$	3.3	2.7	2.1	1.6	1.25
	$f_T \times BV_{CEO}$ (GHz-V)	$\lambda^{-0.45}$	510	570	650	755	865

as evidenced by our cryogenic work,  $f_{max}$  benefits from the less aggressive collector scaling because of lower  $C_{BC}$ , imposing less stress on reducing  $R_B$  to enhance  $f_{max}$  at the same rate as  $f_T$ . In fact, one of the major tasks in the state-of-the-art SiGe HBTs is to minimize  $C_{BC}$ , for both its intrinsic and extrinsic contributions. Newly proposed structures, as mentioned in [173] and [174], rely on advanced Si processes and use STI and self-aligned techniques to reduce the extrinsic  $C_{BC}$  significantly. It is likely that such techniques will to be necessarily employed in the future mainstream SiGe device designs. Therefore the decrease of  $C_{BC}$  will become the major driver for improved  $f_{max}$ , as shown in Table 5.

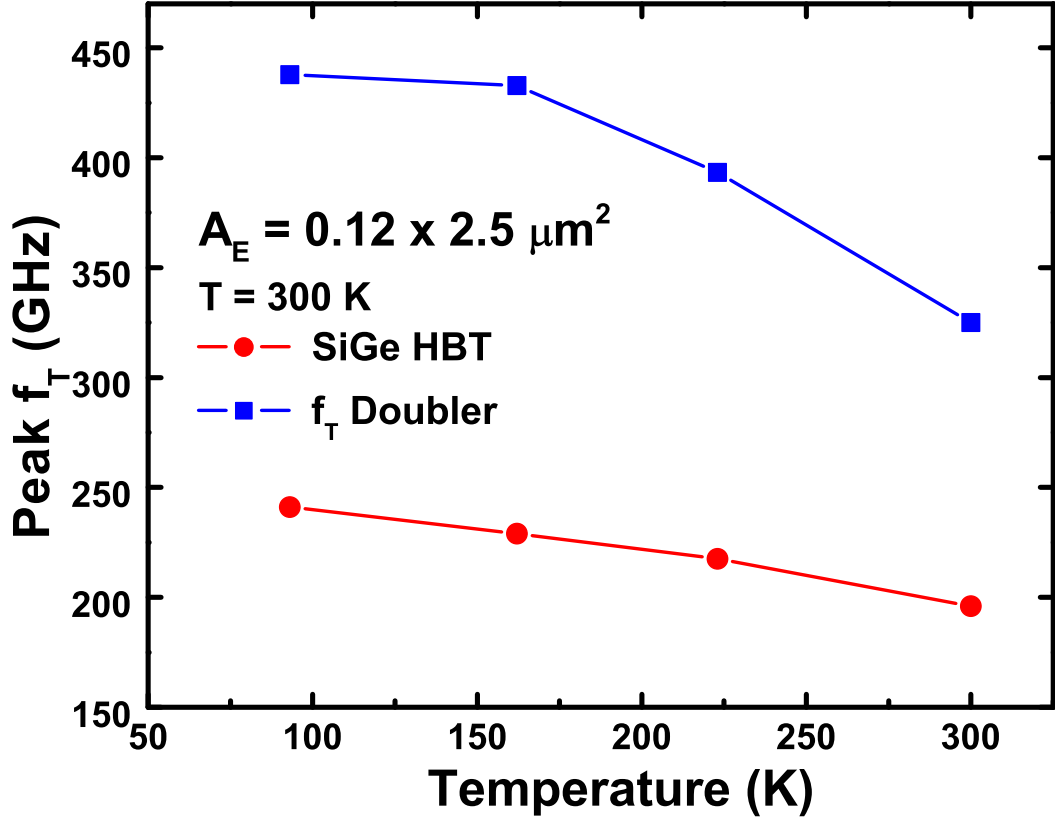
Our cryogenic work in the previous chapter also provides another useful approach



**Figure 66:** Measured peak  $f_T$  as a function of emitter length ( $L_E$ ) for SiGe HBTs and  $f_T$ -doublers.

to reach higher  $BV_{CEO}$ , that is, the lower  $\beta$  at low injection can improve  $BV_{CEO}$ , an important concept for the development of the next generation SiGe HBTs. One can adjust the base and emitter doping profiles and the Ge gradient in the BE junction to reach lower  $\beta$  at low injection. Moreover, the newly developed fully-silicided emitter serves this purpose as well [175]. The fully-silicided emitter also decreases  $R_E$ . As a result, the flare-shaped emitter that is widely used in poly-emitter SiGe HBTs (to decrease the contact resistance) will no longer be necessary. This will improve the scalability of SiGe HBTs, and the potential problem of higher extrinsic  $C_{BE}$  for future SiGe HBTs, as discussed in [90], can be suppressed significantly.

To reach a higher speed with such a conservative collector scaling, as predicted in



**Figure 67:** Measured peak  $f_T$  as a function of temperature for the SiGe HBT and the  $f_T$ -doubler.

Table 5, engineering the BE junction becomes mandatory. This work was historically ignored because of the difficulty in decoupling the  $\tau_{BE}$  from  $\tau_B$  and  $\tau_E$ . Our cryogenic work investigates the importance of  $\tau_{BE}$  by using cooling to effectively adjust the BE junction width and the Ge profile inside the BE junction. However, little work at room temperature has been reported focusing on this part of the device optimization, and [176] shows encouraging evidence of reducing the charging time of  $\tau_{BE}$  without necessarily increasing collector current density. It can be expected that such a reduction of  $\tau_{BE}$  without aggressive collector scaling will be of paramount importance in future scaling of SiGe HBTs in following the proposed roadmap.

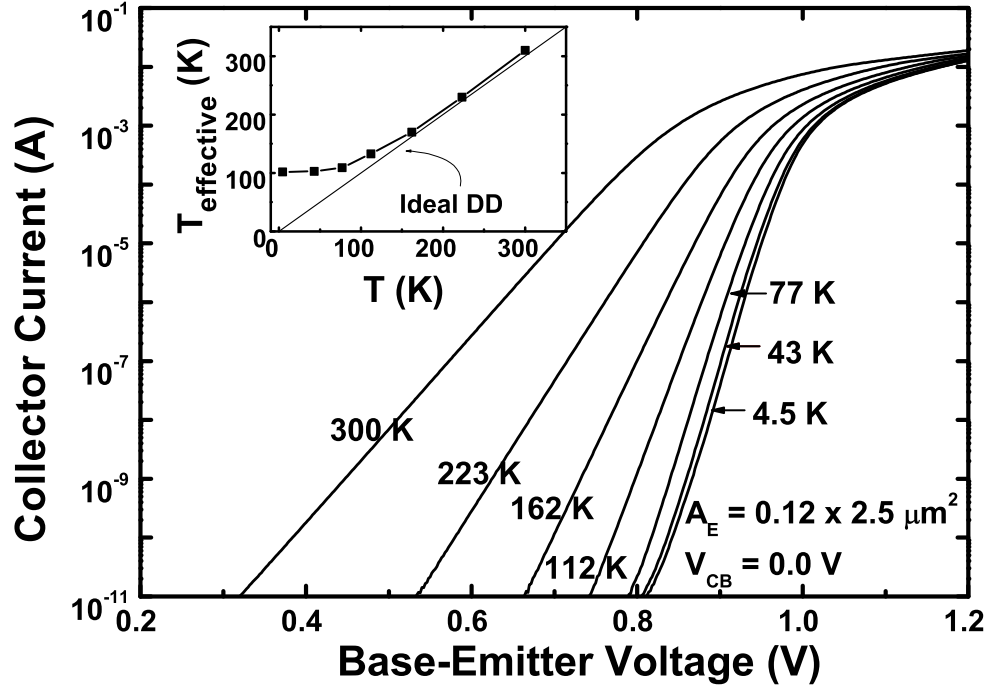
### 5.3 *New Device Physics For THz SiGe HBTs*

Here we discuss non-equilibrium base transport and base recombination (and the associated negative differential resistance (NDR) effect) that may potentially be evident in highly-scaled SiGe HBTs at room temperature, using our cryogenic results as a starting point.

#### 5.3.1 Non-Equilibrium Base Transport

Fig. 68 shows the  $I_C$  at different ambient temperatures for the prototype #3 (mentioned in the previous chapter) fourth-generation SiGe HBT. As classical drift-diffusion theory predicts, the slope increases and the curve moves to the right as the temperature decreases. However, this trend tends to saturate for this device below about 100 K, suggesting that non-equilibrium transport is playing a role in this temperature range. The extracted effective electron temperature as a function of ambient temperature clearly shows this saturation effect (inset of Fig. 68). Consequently, the peak  $\beta$  reaches a maximum of 7,900 at 77 K and then remains nearly constant to lower temperatures;  $g_m$  shows a similar behavior (Fig. 69), suggesting a stronger influence of non-equilibrium transport at very low temperatures compared to that found in wider-based SiGe HBTs [94].

Note that the reverse-Early effect together with BE heterojunction ([39]) also causes non-ideal slope in  $I_C$ , and it is difficult to decouple this effect from non-equilibrium transport from simple cryogenic measurements. The magnitude of the reverse-Early effect, nevertheless, should be a weaker function of generational scaling. Therefore, it is instructional to understand the dependence of non-equilibrium transport on the specific technology scaling node by comparing first-, third-, and fourth-generation SiGe HBTs (prototype #3) at an ambient temperature of 43 K. The inferred effective electron temperatures are 52, 54, and 102 K, respectively, increasing monotonically with vertical scaling node, and almost doubling between third-

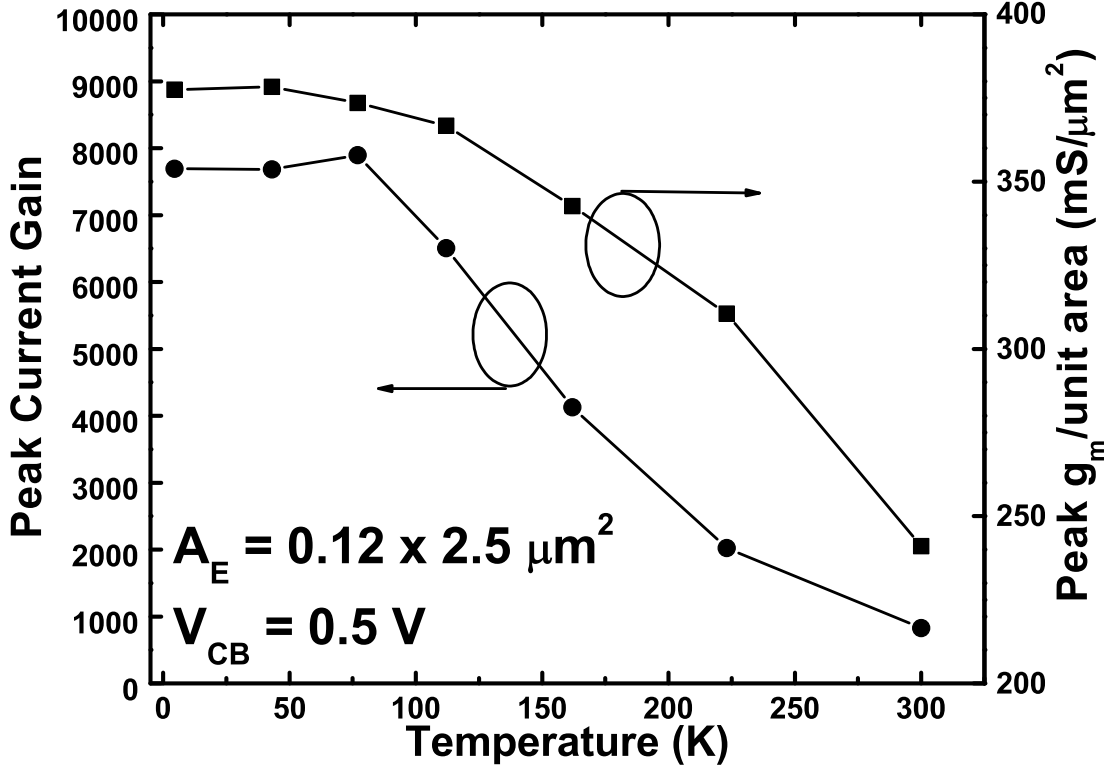


**Figure 68:** Collector current of a fourth-generation prototype #3 SiGe HBT at different temperatures. Inset shows the effective carrier temperature extracted from the slope of  $I_C$  as a function of chuck temperature.

and fourth-generation transistors, indicating the increased role for non-equilibrium effects at the limits of vertical scaling in SiGe HBTs. Comparing the present results with the third-generation SiGe HBT at the same lithography node and similar current density allows us to exclude the influence of other secondary mechanisms in this result. It can be expected that non-equilibrium transport will be significant in the future scaled SiGe HBTs at room temperature, where aggressive vertical scaling further accentuates this non-classic “quasi-ballistic” base transport.

### 5.3.2 Base Recombination and NDR

Base recombination current can be expected to contribute more to the total base current in highly-scaled SiGe HBTs because: 1) the carbon in the base can potentially



**Figure 69:** Peak transconductance and peak current gain at  $V_{CB} = 0.5$  V for a fourth-generation prototype #3 SiGe HBT as a function of temperature.

act as a trap, enhancing Shockley-Read-Hall (SRH) recombination; and 2) higher  $\beta$  reduces the traditional diffusive base current component. This recombination effect can be enhanced by cooling as well as by technology scaling, and will be shown experimentally below.

The base current of the conventional well-made SiGe HBTs is mainly composed of a diffusion component, and the slope of  $I_B$  in the Gummel characteristics should also demonstrate  $\frac{\partial I_B}{\partial V_{BE}} \simeq \frac{qI_B}{kT}$  or  $\frac{\partial(\ln I_B)}{\partial V_{BE}} \simeq \frac{q}{kT}$  at low and medium injection levels. However, such  $V_{BE}$  dependence is no longer valid for the prototype fourth-generation SiGe HBTs at low temperatures, as shown in Fig. 41 in Chapter 4, where the slope of  $I_B$  is shown to be insensitive to temperature. The high  $\beta$  of these devices at low temperatures dramatically decreases the traditional diffusion component of the total

current, while recombination centers in the base significantly enhances recombination, giving rise to an extra base current component ( $\Delta I_B$ ) that is only weakly temperature dependent.

The direct result of this enhancement of  $\Delta I_B$  is the negative differential resistance effect under forced- $I_B$  input drive, as shown in Fig. 43 in Chapter 4. This effect is more evident at higher injection levels, but is not observable in the forced- $V_{BE}$  output characteristics, as shown in the right figure of Fig. 23 in Chapter 2. The heterojunction barrier at fixed  $I_C$  decreases as  $V_{CE}$  increases, resulting in a sudden increase in  $\Delta I_B$ . Consequently,  $V_{BE}$  necessarily shifts to a smaller value to maintain the same  $I_B$  in the forced- $I_B$  measurement,  $I_C$  drops, and the NDR is induced. A related hysteresis effect can also be observed. This unique cryogenic phenomena is enhanced with technology scaling, which is fundamentally different from other NDR effects induced by the enhanced positive feedback due to heterojunction barrier effect (Fig. 23). The latter is more often observed in the first-generation SiGe HBTs, and becomes less important with device scaling. More detailed discussion of these phenomena can be found in Chapter 2.

We expect that this non-ideal base current component may play a role in further scaling of SiGe HBTs for THz speeds (at room temperature). The induced NDR and hysteresis effects are not only potentially important for circuits, but will likely also be challenging for accurate compact modeling using standard models (e.g., with VBIC, HICUM or MEXTRAM).

#### ***5.4 TCAD Study of High-Speed SiGe HBTs***

The DESSIS 10 TCAD tool was used in this study [177], with the hydrodynamic model turned-on, together with the energy balance equations that include both the energy relaxation and the collision terms. Carrier mobilities were captured by the Philips unified mobility model, while the high field saturation was calculated using



the Canali model by using carrier temperatures as the driving force. In order to accurately model avalanche behavior in SiGe HBTs, the ionization coefficient was calculated through the Okuto-Crowell model, with the driving force derived from the local carrier temperatures. This model and its implications on improving breakdown voltages will be further discussed in the next section.

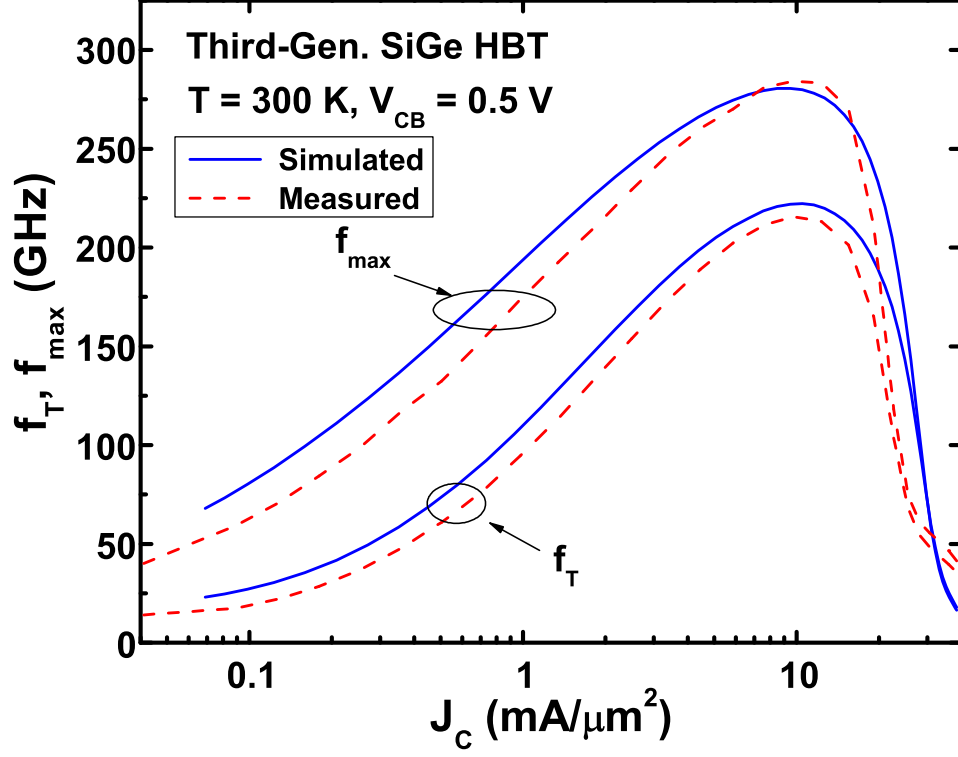
#### 5.4.1 TCAD Calibration

The model parameters used in TCAD were carefully calibrated on a third-generation SiGe HBT (IBM 8HP). Figs. 70 and 71 show that the simulated  $f_T$ ,  $f_{max}$ , and multiplication factor ( $M - 1$ ) match well with the measured data, verifying the validity of the physical models used in the TCAD simulations.

The emitter recombination velocity was tuned to calibrate the Gummel characteristics and  $\beta$  (peak  $\beta = 500$  for this technology). After  $\beta$  is calibrated, both the base current ( $I_B$ ) reversal point and the bias point where  $\beta(M - 1) = 1$  can be used to calculate  $BV_{CEO}$ . For example, in Fig. 71,  $I_B$  reverses sign at  $V_{CB} = 1.1$  V, where  $M - 1 = 0.002$ . Given that  $V_{BE} = 0.7$  V and  $\beta = 500$ , both methods produce a consistent value of  $BV_{CEO} = 1.8$  V.

#### 5.4.2 Scaling Examples

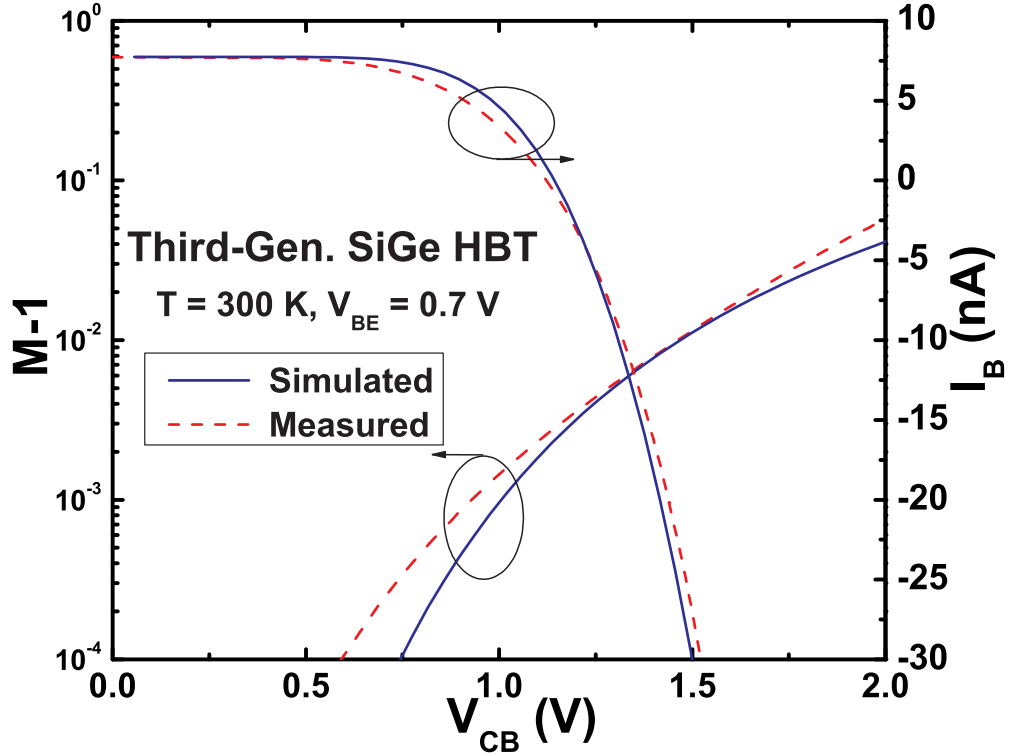
Most of the targeted performance in Table 5 can be achieved in almost every technology node. Table 6 lists the key device parameters of IBM 8HP (130 nm node) and those SiGe HBTs simulated at Georgia Tech (90 to 45 nm node). Tables 5 and 6 are consistent all the way to the 610 GHz / 750 GHz of  $f_T$  /  $f_{max}$  at 45 nm lithography node. However, further lateral scaling results in an unexpectedly stronger peripheral effects that require significantly higher current density to compensate for. To avoid such an unrealistically high emitter current density (about  $100\text{mA}/\mu^2$ ),  $W_E = 32$  nm is discarded in the roadmap. Instead, we stick to the 45 nm emitter window and optimized the extrinsic base region to make up for the high intrinsic base



**Figure 70:** Comparison of measured and simulated  $f_T$  and  $f_{max}$  vs.  $J_C$  of a third-generation SiGe HBT.

resistance associated with the wider emitter window. The final device is labeled as “GT 45 nm #2”.

As shown in Table 6, base width ( $W_B$ ) scaling is proportional to the total transit time. Consequently,  $\tau_B$  maintains a near-constant portion of the total  $\tau_{EC}$ . Note that even using drift-diffusion theory,  $\tau_B \propto W_B$  in an ultra-thin base region where the exit velocity term dominates [56]. The increase of the collector doping level ( $N_C$ ) sustains the increasing  $J_C$  at peak  $f_T$  (or  $J_{C,peak}$ ) in Table 5, as  $J_{C,peak}$  follows the roadmap at 90, 65, and 45 nm lithography node, as shown in Figs 72 and 73. The roadmap, however, underestimates the  $J_{C,peak}$  by 20% in GT 45 nm #2 as compared to the 32 nm node in Table 5. Although high  $J_C$  is normally desirable for better



**Figure 71:** Comparison of measured and simulated  $M - 1$  and  $I_B$  vs.  $V_{CB}$  of a third-generation SiGe HBT.

linearity, it nonetheless raises serious reliability issues regarding the silicided emitter contact with further device scaling. The current  $J_{C,peak}$  by 20% in “GT 45 nm #2” is sustainable in current IBM technology; however, this will become the single most important bottleneck for THz scaling of SiGe HBTs. Apparently,  $f_T$  of 790 GHz and  $f_{max}$  of 960 GHz are achievable in the 45 nm technology node.

### 5.5 Superjunction Collector Design for Improving Breakdown Voltage

In this section, we propose a novel idea for using a “superjunction collector” to significantly improve the avalanche breakdown behavior of SiGe HBTs with little or no impact on device speed, and apply it to the design of a SiGe HBT with a target

**Table 6:** Key device parameters of IBM 8HP (130 nm) and SiGe HBTs simulated at Georgia Tech (90 to 45 nm).

$W_E$	130 nm	90 nm	65 nm	45 nm #1	45nm #2
Peak $f_T$ (GHz)	207	375	442	614	789
Peak $f_{max}$ (GHz)	285	445	545	750	958
Base Width ( $W_B$ )	$W_{B0}$	$0.63W_{B0}$	$0.45W_{B0}$	$0.34W_{B0}$	$0.27W_{B0}$
Base Doping ( $N_B$ )	$N_{B0}$	$1.9N_{B0}$	$1.5N_{B0}$	$1.5N_{B0}$	$1.5N_{B0}$
$R_{BI}$ ( $k\Omega/\text{square}$ )	2.0	1.7	3.0	4.0	5.0
Collector Doping ( $N_C$ )	$N_{C0}$	$2.4N_{C0}$	$3.6N_{C0}$	$3.6N_{C0}$	$7.2N_{C0}$

performance of 100 GHz peak  $f_T$  and 3.0 V  $BV_{CEO}$ .

### 5.5.1 Abrupt Superjunction Design

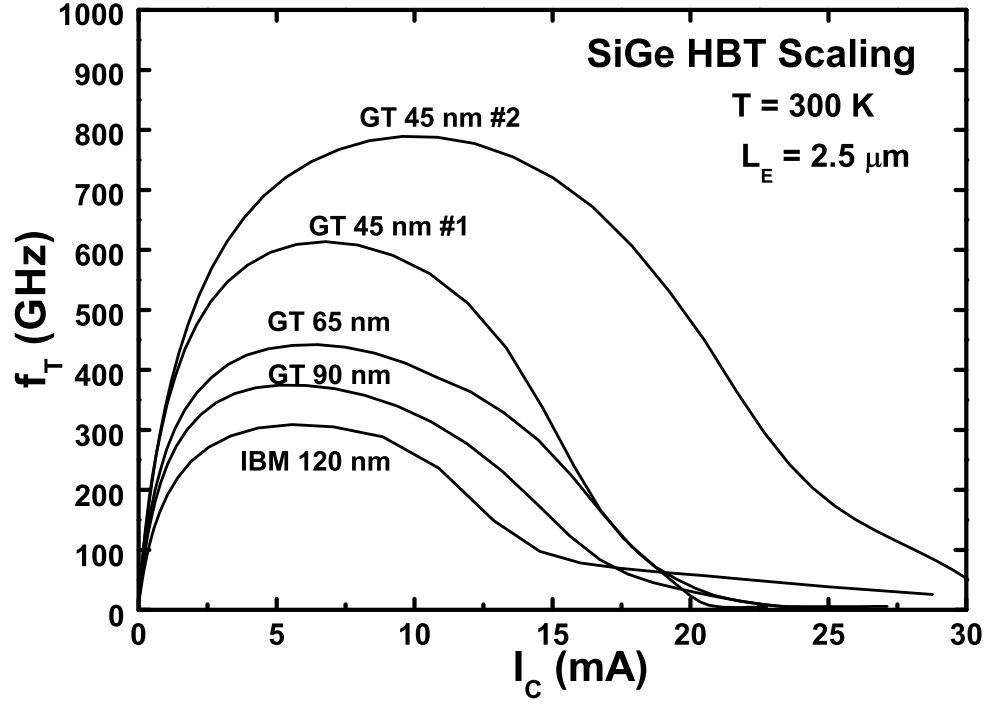
As we have discussed, a lower  $\beta$  can always lead to a higher  $BV_{CEO}$ . In order to maintain a useful DC current gain, all simulated devices in this work maintain  $\beta$  of above 200.

Based on the third-generation SiGe HBT device structure, the highest  $BV_{CEO}$  that can be obtained by standard SIC optimization is 2.67 V. This is achieved by lowering the SIC doping level of third-generation SiGe HBTs. Doing so lowers  $f_T$  and  $J_C$  at the peak  $f_T$  bias ( $J_{C,peak}$ ) while improving  $f_{max}$  and the avalanche behavior of the CB junction.

The non-local avalanche behavior in the standard collector optimization of a SiGe HBT is shown in Fig. 74. This figure demonstrates a “dead space region” (as defined in [39]) between the peaks of the electron temperature ( $T_n$ ) and the electric field. A careful study of this non-local effect suggests that proper tuning of the electric field in the CB SCR could potentially improve the avalanche breakdown behavior.

As defined in the Okuto-Crowell model in [177], the electron ionization coefficient ( $\alpha_n$ ) can be expressed as:

$$\alpha(E_n^{eff}) = a \cdot (1 + c(T_n - T_0)) \cdot E_n^{eff} \cdot \left( \frac{b[1 + d(T_n - T_0)]}{E_n^{eff}} \right)^\delta, \quad (27)$$



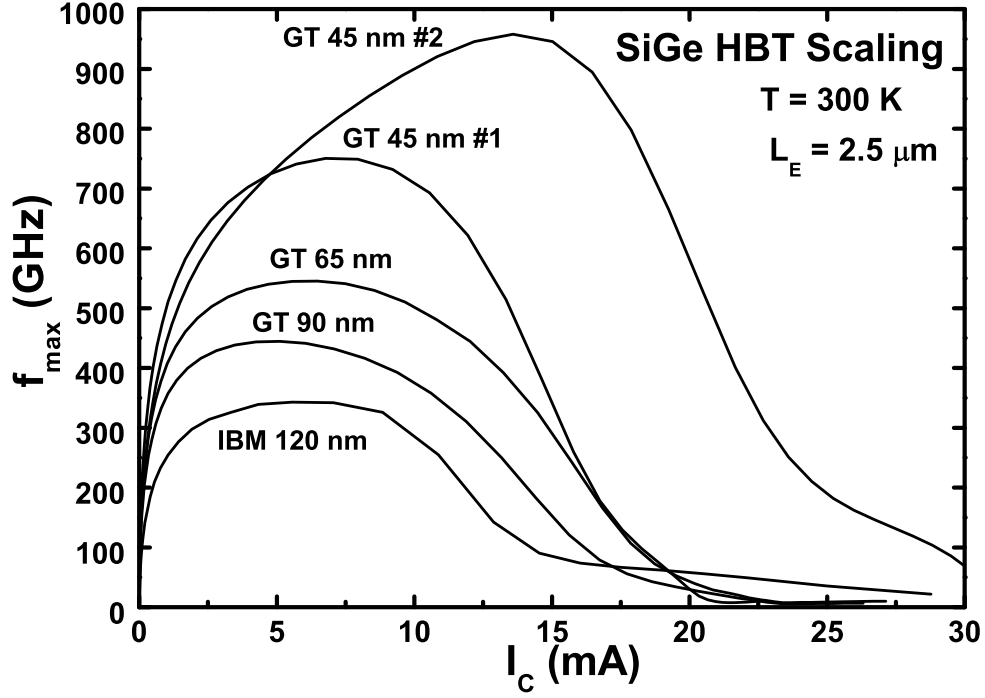
**Figure 72:** Cutoff frequency ( $f_T$ ) as a function of  $I_C$  for a fourth-generation SiGe HBT and four experimental SiGe HBTs simulated at Georgia Tech.

where  $a$ ,  $b$ ,  $c$ ,  $d$ , and are fitting parameters. Non-local effects are taken into account by using  $T_n$  to calculate the effective field,  $E_n^{eff}$ , as described in [177]

As an indication of the electron kinetic energy,  $T_n$  affects  $\alpha_n$  exponentially, as shown in Eq. 27. An exponential increase in  $\alpha_n$ , in turn, dramatically increases the electron avalanche generation rate ( $G^{ii}$ ), which can be approximated by:

$$G^{ii} = \alpha_n \cdot n \cdot v_n. \quad (28)$$

The high electron kinetic energy near peak  $T_n$  is mainly obtained by electrons traveling

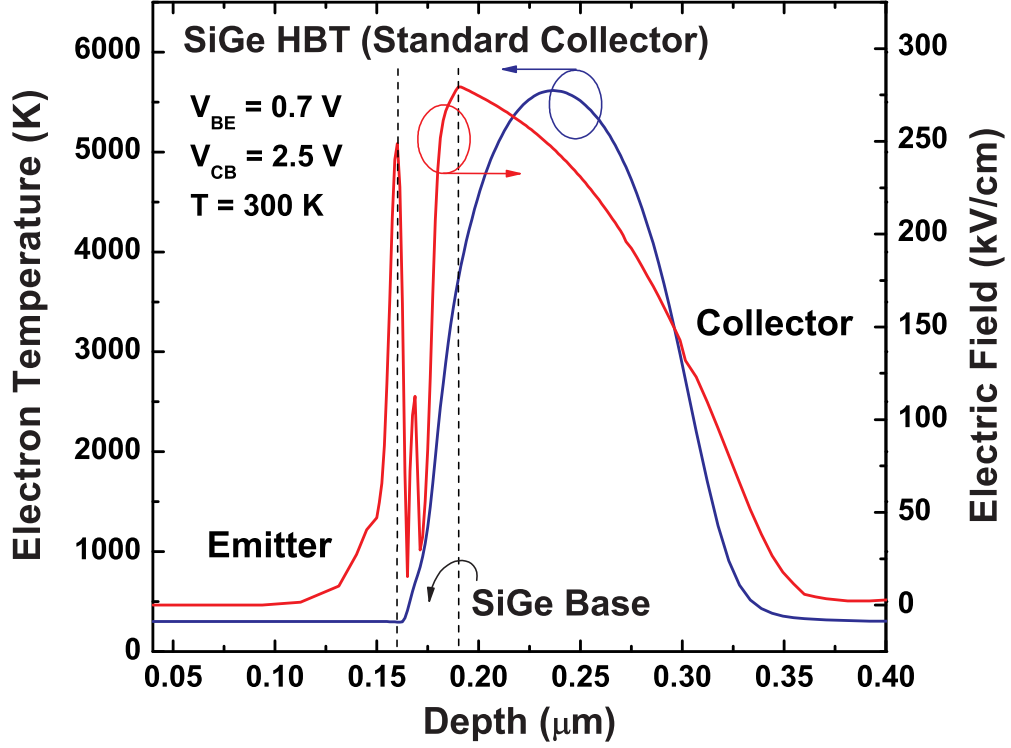


**Figure 73:** Maximum oscillation frequency ( $f_{max}$ ) as a function of  $I_C$  for a fourth-generation SiGe HBT and four experimental SiGe HBTs simulated at Georgia Tech.

through the “dead space region” where the high electric field rapidly accelerates the electrons. If part of the high electric field could somehow be shifted deeper into the CB SCR,  $T_n$  could be significantly reduced while maintaining the same voltage drop across the SCR, leading to a lower avalanche current at the same  $V_{CB}$ .

In terms of AC performance, the transit time through the CB SCR is mainly determined by the width of the depletion region due to electron velocity saturation [56]. Standard SIC optimization for higher  $BV_{CEO}$  decreases the collector doping level, which inevitably widens the CB SCR, giving rise to a higher CB transit time and degrading  $f_T$ .

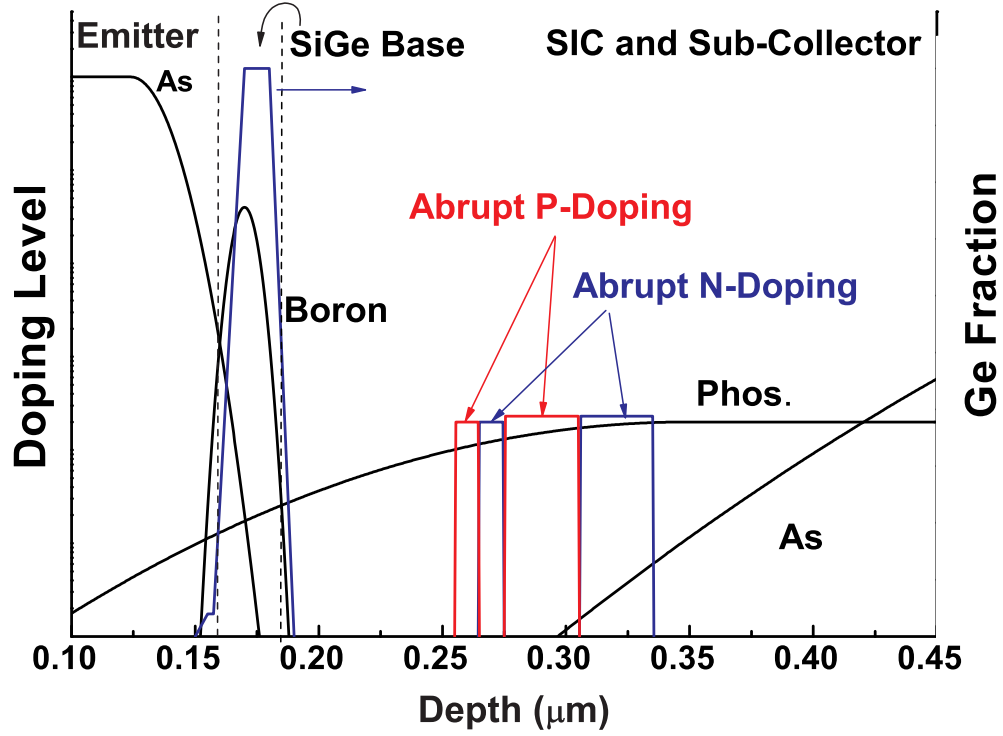
Therefore, a “superjunction” technique is proposed here as a means to lower the peak CB electric field. As shown in Fig. 75, two abrupt PN junctions (a superjunction)



**Figure 74:** Simulated electron temperature and electric field in a SiGe HBT with standard collector (the peak  $f_T = 105$  GHz and  $BV_{CEO} = 2.67$  V ).

were carefully designed and placed into the CB SCR. The electric field in the depletion region can be favorably adjusted in this way. In addition, since no net total charge is put into the CB depletion region, (the p and n-doping cancel out), the superjunction does not change the width of the CB SCR, and the AC performance of the transistor is not compromised.

TCAD simulations were conducted on SiGe HBTs with a standard collector design and a superjunction collector design, both shown in Fig. 75. Besides the four alternatively-doped layers that form the superjunction, all other doping profiles are the same for the two devices for ease of comparison. The simulation results are shown in Figs. 76-79.

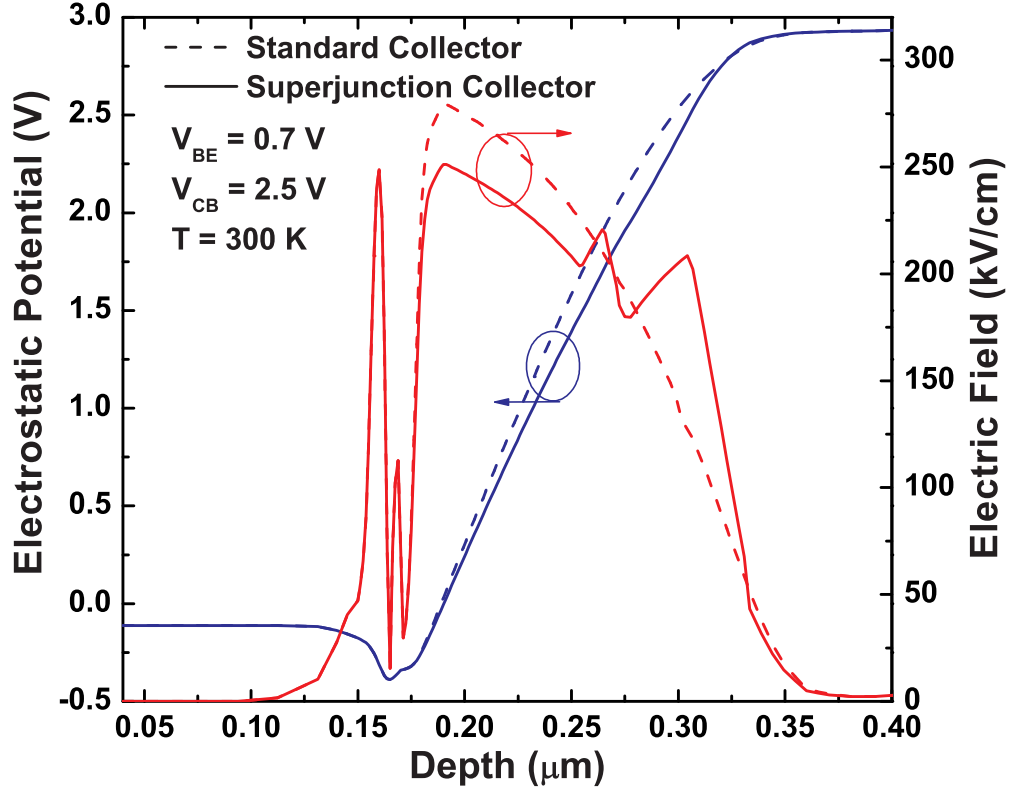


**Figure 75:** The doping profile used in the TCAD simulation, including the standard collector doping and the PN superjunction.

The higher electric field induced by the superjunction shifts more voltage drop deeper into CB SCR, as evidenced by the lower electric field near the CB metallurgical junction, and the higher electric field in the deep CB SCR (Figs. 76). Consequently, comparing the electrostatic potentials of the two profiles, one can easily see that the superjunction collector design shows a smaller slope in the electrostatic potential at the shallow CB SCR (mostly in the “dead space region”) and a steeper slope in the deep CB SCR. This is a clear evidence of the electric field shift.

As a result, the lower peak electric field in the “dead space region” decreases the peak  $T_n$  by 10% (Fig. 77). Due to the exponential dependence of the ionization coefficient on  $T_n$ , as described in Equation (27), the peak  $\alpha_n$  is nearly halved in the superjunction collector design (Fig. 77). The resultant  $M-1$  is decreased by almost an

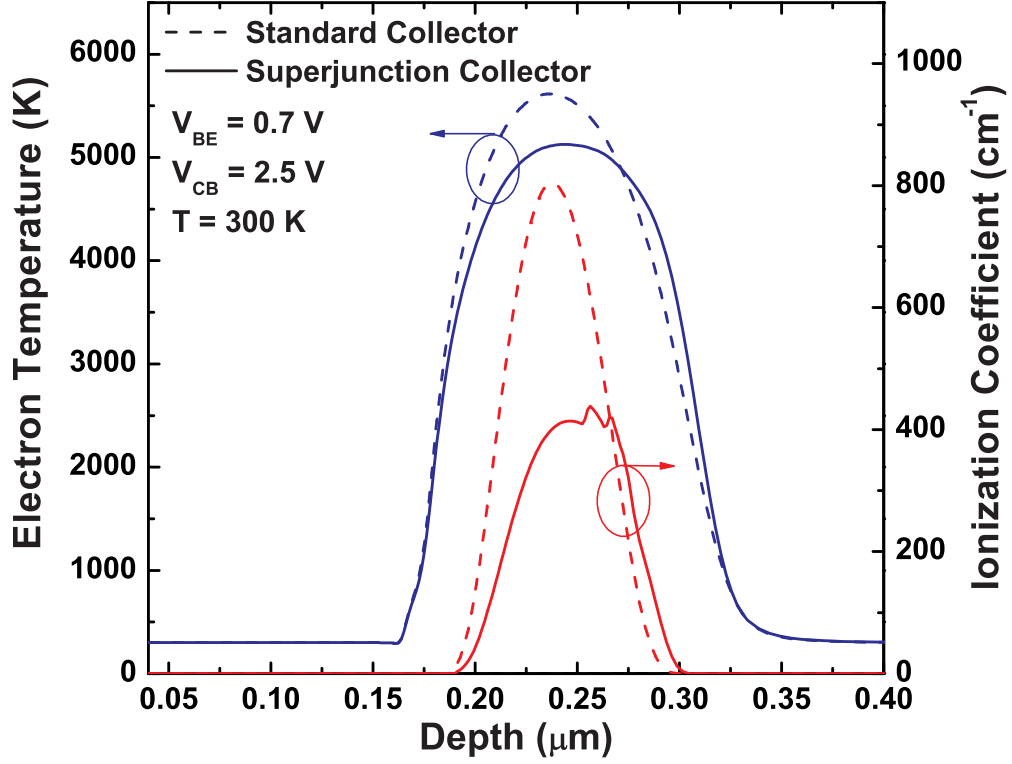




**Figure 76:** Comparison of electrostatic potential and electric field of the standard collector with that of a superjunction collector.

order of magnitude at fixed  $V_{CB}$ , and, therefore,  $BV_{CEO}$  improves by 0.33 V (Fig. 78), a 12% increase over the standard collector design, and clearly of significance to circuit designers.

Since no net total charge is put into the CB depletion region, the superjunction does not change the width of the CB SCR to first order. Consequently, it has little influence on the device speed. This was verified, as shown in Fig. 79, where the minor  $f_T$  degradation ( $< 4\%$ ) is caused by the more severe heterojunction barrier effect in the new collector design owing to the smaller voltage drop near the CB metallurgical junction. This small degradation occurs at peak- $f_T$  current density and above and is



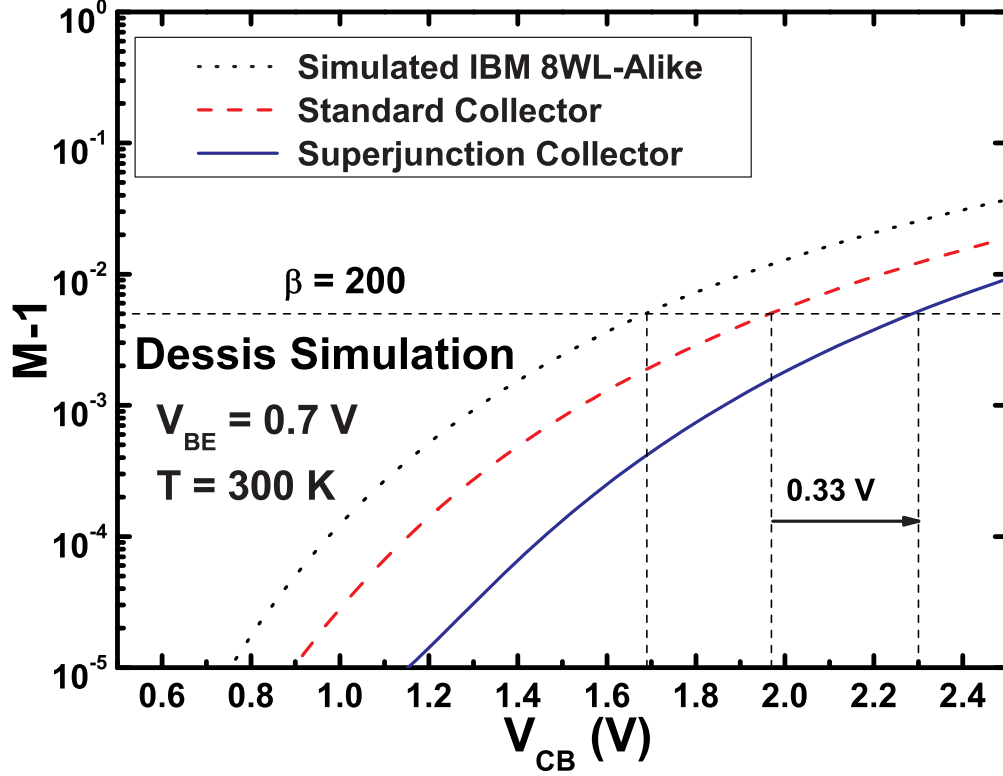
**Figure 77:** Comparison of electron temperature and ionization coefficient of the standard collector with that of a superjunction collector.

thus not important for most circuit applications. In addition, with higher operating voltage, the higher  $V_{CB}$  of the new design decreases the CB capacitance, which in turn enhances peak  $f_{max}$  by about 10 GHz (Fig. 79).

### 5.5.2 Practical Profile Design

Thermal cycles have little impact on the superjunction device performance. Fig. 80 shows the Gaussian distribution of the PN superjunction after the doping in Fig. 75 goes through thermal cycles. The resultant device still shows a peak  $f_T / f_{max}$  of 100 / 349 GHz at a  $BV_{CEO}$  of 3.02 V (Fig. 81).

A more simplified doping profile can be achieved if we merge the two phosphorous

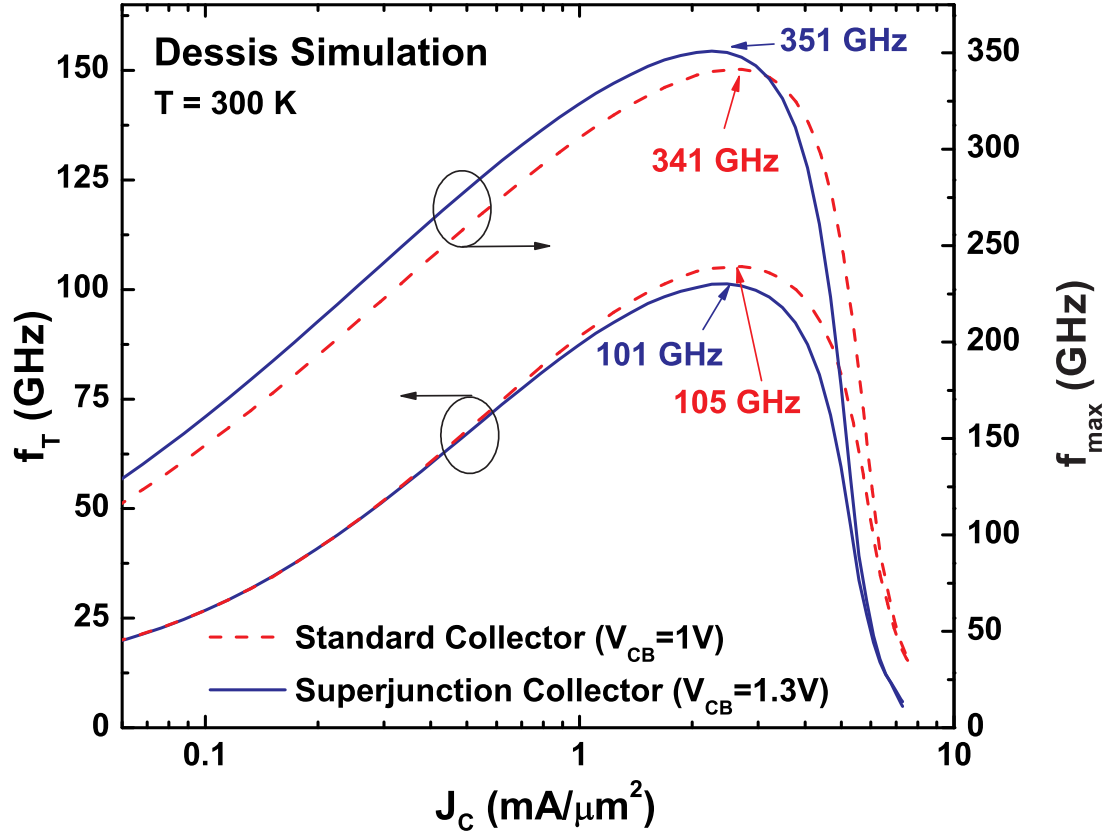


**Figure 78:** Simulated multiplication factor for the IBM 8WL, standard collector, and superjunction collector SiGe HBTs.

implants into the background SIC doping (Fig. 82). In this way, only two Boron implants are required, and only one block-out mask is needed.

The simplified superjunction collector SiGe HBT shows an ideal output characteristics (Fig. 83), improved avalanche behavior (Fig. 84), and ideal AC performance with  $f_T = 105$  GHz and  $f_{max} = 351$  GHz (Fig. 85).

Nevertheless, the added superjunction in the CB SCR affects device speed at a low  $V_{CB}$ . This is because the CB SCR decreases its width at lower CB biases, and the superjunction prevents SCR width from shrinking at low  $V_{CB}$ , giving rise to a higher transit time. As a result, one should expect degraded frequency response when the

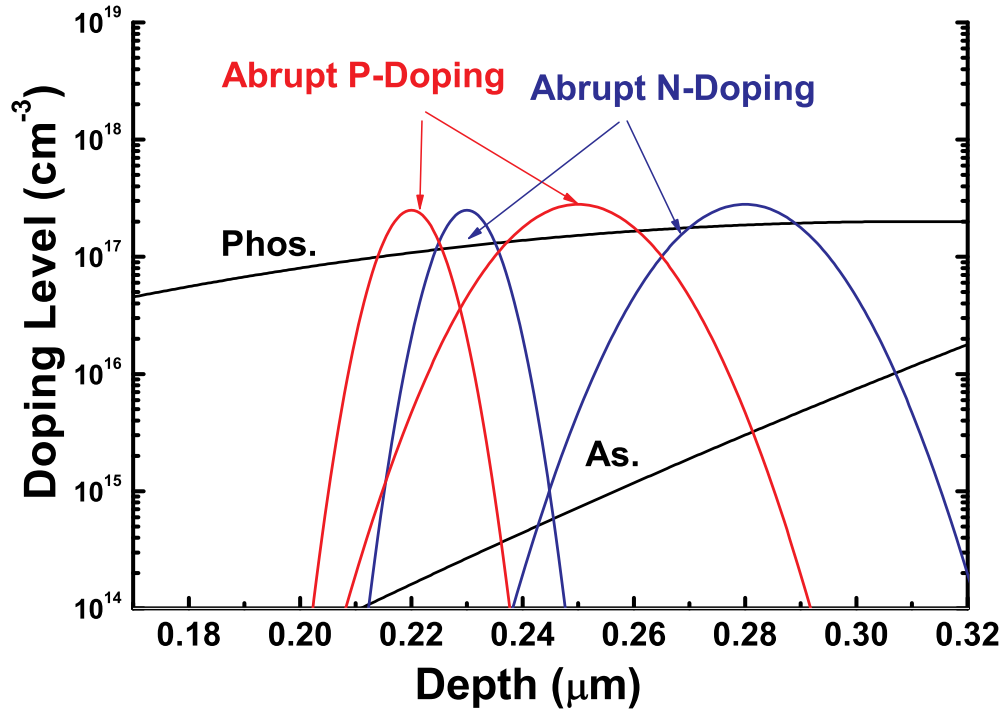


**Figure 79:** Comparison of  $f_T$  and  $f_{max}$  of the standard collector SiGe HBT with that of a superjunction one.

superjunction device is biased in saturation. This effect can indeed be observed in Fig. 86. The solution to this problem is to not place the superjunction too deep into the CB SCR. As shallow superjunctions cannot fully utilize the dead space effect, the enhancement of the breakdown voltage will be affected. Therefore, a trade-off exists between the low- $V_{CB}$   $f_T$  and  $BV_{CEO}$  improvement, and our simplified superjunction in Fig. 82 presents a good balance between these two, as evidenced by Figs. 84 and 86.

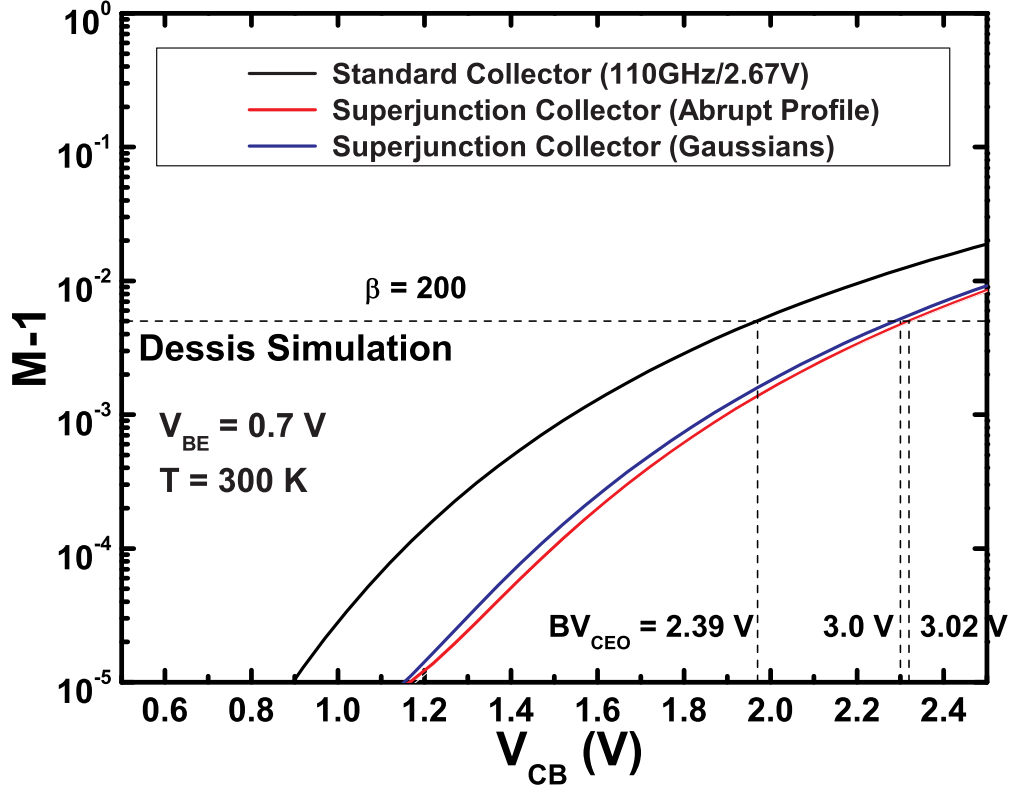
### 5.5.3 Benchmarking

In the bipolar transistor design, there is a well-known fundamental trade-off between  $f_T$  and the breakdown voltage ( $BV_{CEO}$  or  $BV_{CBO}$ ). While  $BV_{CEO}$  is the result of a



**Figure 80:** The doping profile of Gaussian superjunction representing the impact of thermal cycles.

positive feedback mechanism and can be increased by either decreasing the current gain or improving the avalanche behavior of the CB junction,  $BV_{CBO}$  can only be enhanced by improved CB junction designs. In many RF circuit designs, both  $BV_{CEO}$  and  $BV_{CBO}$  are important, since together they set the maximum operating voltages in common-emitter mode, and common-base mode, respectively. Clearly, the increase in  $BV_{CEO}$  in this work is not caused by  $\beta$ . Indeed the superjunction collector design improves the avalanche behavior of the CB junction, which should improve both  $BV_{CEO}$  and  $BV_{CBO}$ . However, the improved  $BV_{CBO}$ , which is usually several volts higher than  $BV_{CEO}$ , is not reported in this work, due to the limitation of the calibrated Okuto-Crowell model used in TCAD. An actual fabricated SiGe HBT is needed for an accurate study of the  $BV_{CBO}$  of the “superjunction collector” device, and this

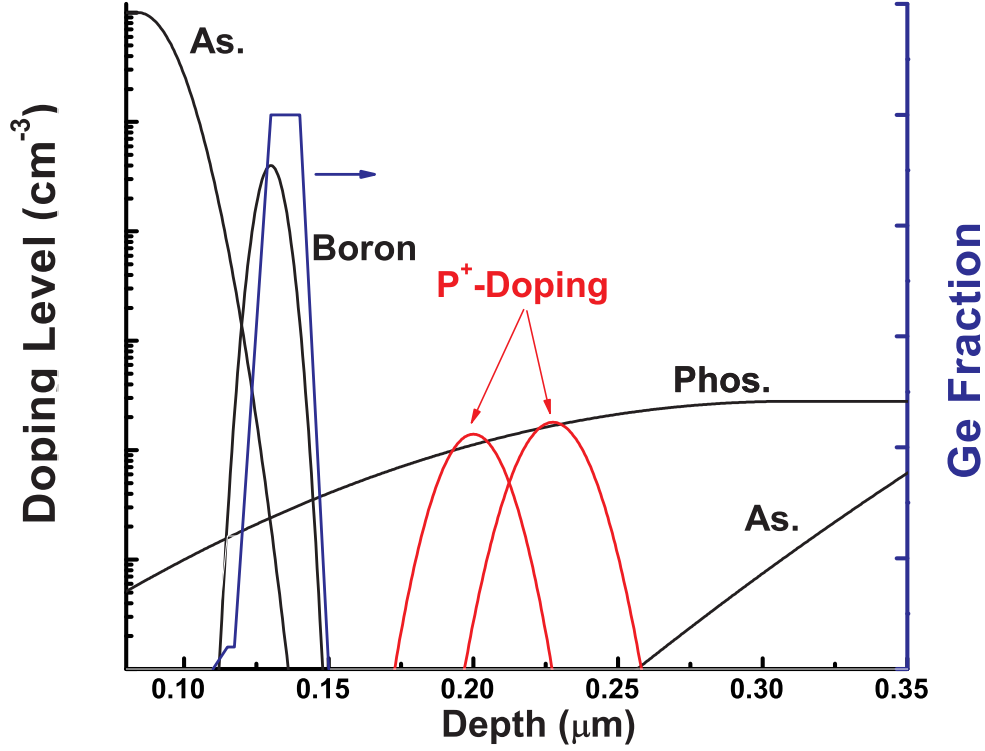


**Figure 81:** Simulated multiplication factor for a standard, abrupt superjunction, and Gaussian superjunction collector SiGe HBTs.

experimental work is in progress.

Fig. 87 shows the extent to which this novel superjunction collector SiGe HBT can improve the state-of-the-art. The curved line indicates the trade-off between  $f_T$  and  $BV_{CEO}$ . Conventional profile optimization of the SIC doping cannot move above this line; however, the new superjunction collector can break this barrier, achieving a device with:  $f_T = 101$  GHz,  $f_{max} = 351$  GHz, and  $BV_{CEO} = 3.0$  V (at 130 nm).

We believe that such a novel collector design can either be fabricated using a fully epitaxially-grown structure, as demonstrated in [174], or possibly by using ion implantation combined with laser annealing. This superjunction SiGe HBT design

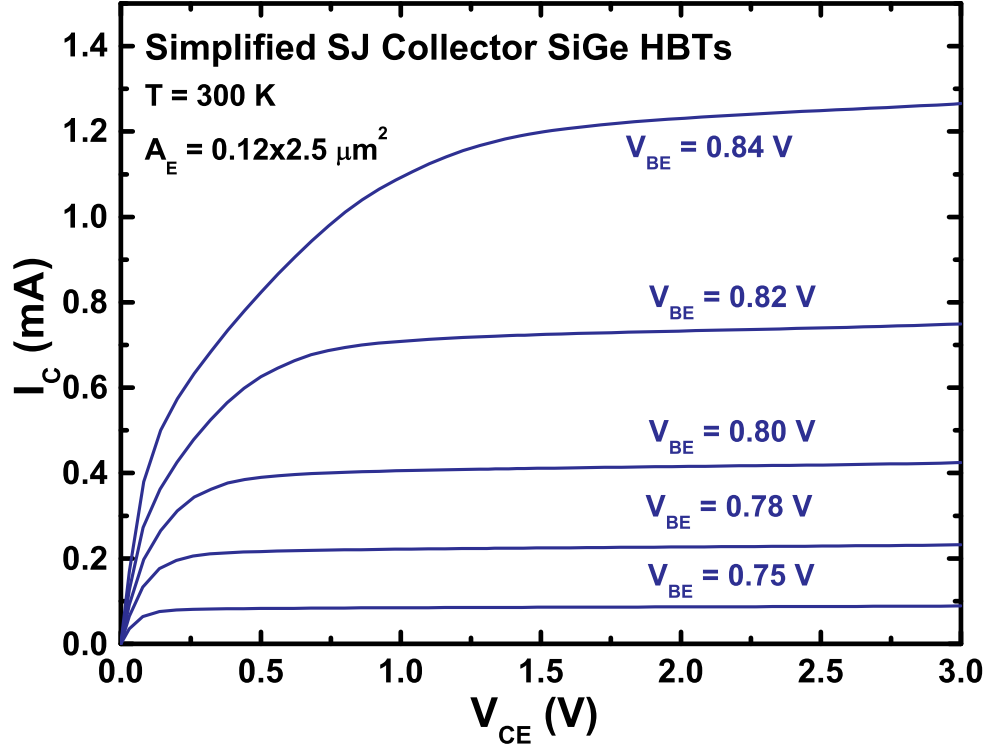


**Figure 82:** The doping profile of the simplified superjunction collector SiGe HBT.

can potentially be applied to all epi-base bipolar transistors (NPN or PNP) to achieve improved breakdown voltage without performance loss.

#### 5.5.4 Comparison with Other Inventions

The structure proposed in the present work is significantly different from that found in [122], [120], and [178]. Superjunction techniques were historically employed in the body or substrate of semiconductor power devices to improve the breakdown voltage / on-state resistance ( $R_{on}$ ) trade-off (e.g., [122]). While enabling a high doping level to reduce  $R_{on}$ , such a superjunction in a high-power device maintains an SCR that is wide enough to sustain a bias of more than 50 V. However, the superjunction described here for use in high-speed SiGe HBTs is aimed at altering the  $T_n$  profile

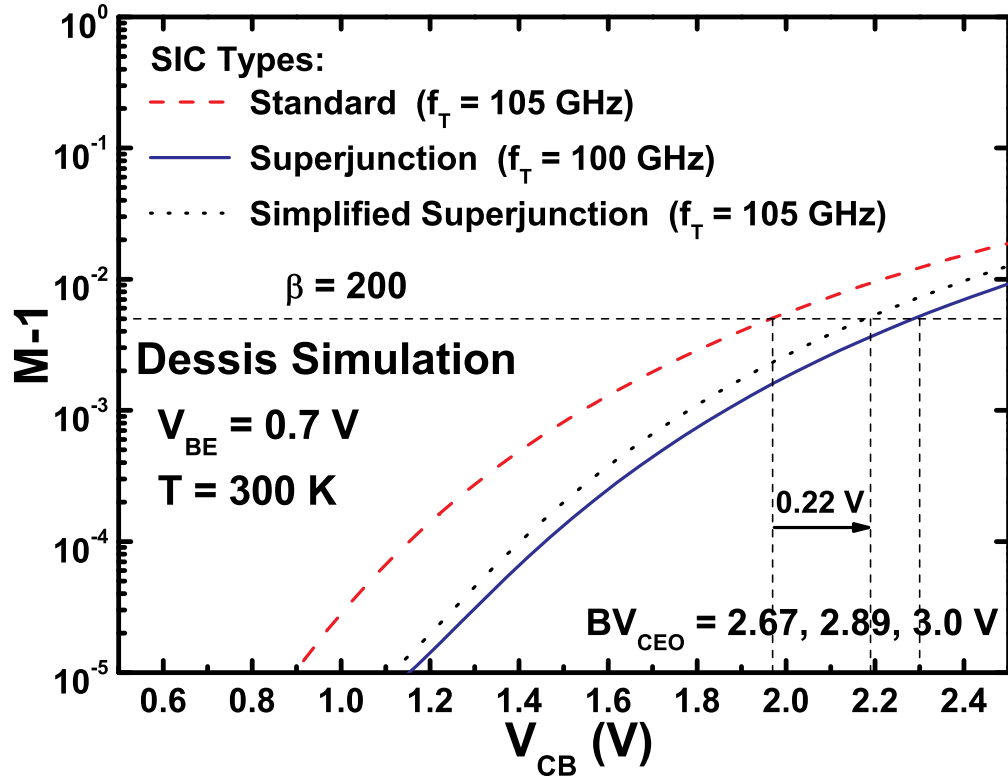


**Figure 83:** Output characteristics of the simplified superjunction collector SiGe HBT.

while preserving the minimum CB SCR width, and thus maintaining high speed operation.

In [120] and [178], an  $i\text{-p}^+$  layer was inserted between the base and collector of a GaAs HBT to form an inverted field that prevents the electrons from moving out of the  $\Gamma$ -valley. In this way, inter-valley scattering is minimized and near-ballistic transport of electrons can be achieved. Such a structure was designed to utilize velocity overshoot in GaAs, however, and hence, cannot be applied to Si-based bipolar transistors. In fact, such an  $i\text{-p}^+$  layer located between the base and collector regions would degrade the frequency response of a SiGe HBT because of the increased effective base transit time.



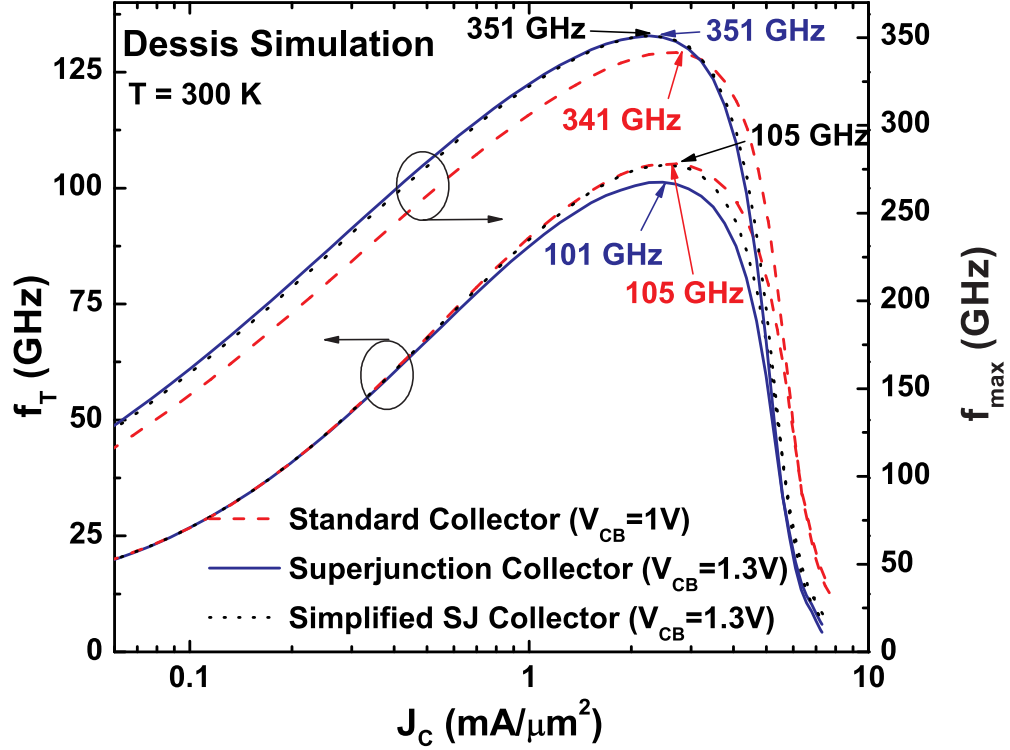


**Figure 84:**  $M - 1$  for a standard, abrupt superjunction, and simplified superjunction collector SiGe HBTs.

## 5.6 Summary

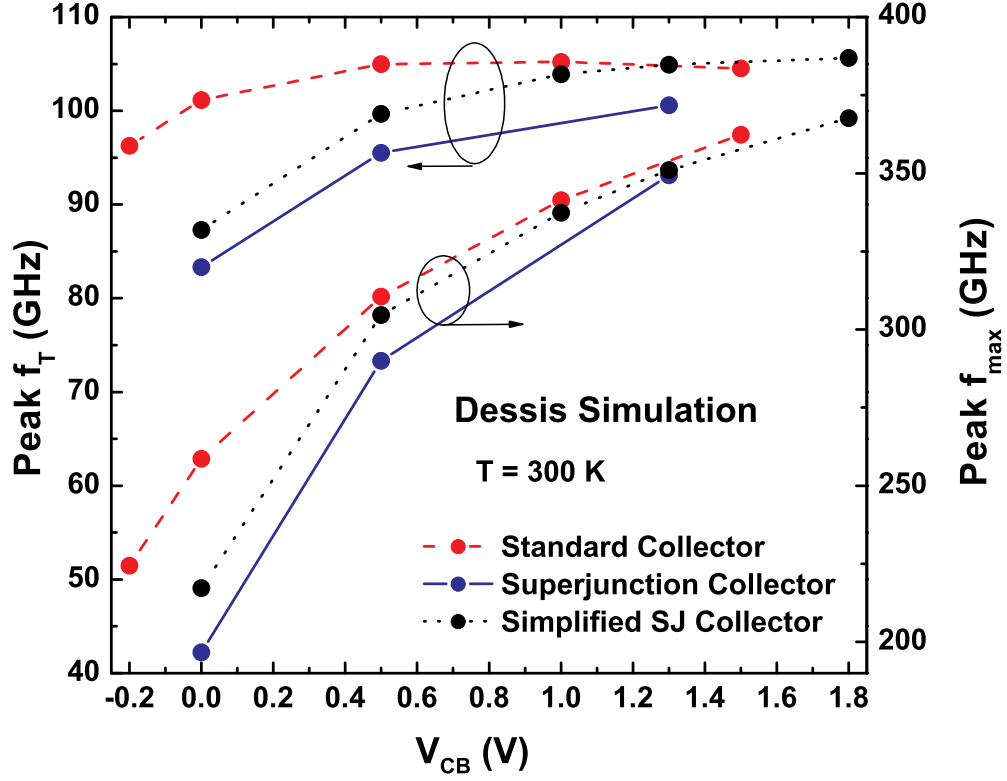
In this chapter, a SiGe HBT  $f_T$ -doubler topology improves the speed of 130-nm, 200-GHz, third-generation SiGe HBTs up to 325 GHz at room temperature and 438 GHz at 93 K. A comprehensive investigation of the small-signal model of the doubler is conducted, for the first time, to explore the origins and implications of the  $f_T$  improvement. It is shown that the  $f_T$ -doubler can be treated as a single transistor unit cell for circuit design. Issues such as emitter geometry scaling and temperature dependence are investigated, further providing useful information for circuit designers.

We have also investigated the path towards achieving terahertz speeds in SiGe



**Figure 85:**  $f_T$  and  $f_{max}$  of a standard, superjunction, and simplified superjunction collector SiGe HBTs.

HBTs by examining how device scaling techniques couple of the cryogenic DC and AC characteristics of these transistors. A new roadmap of SiGe HBT device scaling is given, and the research work on the cryogenic SiGe HBTs helps to identify the key components for future scaling, as well as some interesting device physics that can indeed be observed in these highly-scaled transistors, and which may potentially play a role in sustained scaling for THz performance at 300 K. TCAD simulations with well calibrated hydrodynamic models were used for the scaling study of SiGe HBTs, and it is shown by TCAD simulation that  $f_T$  of 790 GHz and  $f_{max}$  of 960 GHz are achievable in the 45 nm technology node..



**Figure 86:**  $f_T$  vs.  $V_{CB}$  of a standard, superjunction, and simplified superjunction collector SiGe HBTs.

We have also proposed a novel idea for using a “superjunction collector” to significantly improve the avalanche breakdown behavior of SiGe HBTs with little or no impact on device speeds, and apply it to the design of a SiGe HBT with 100 GHz peak  $f_T$  and 3.0 V  $BV_{CEO}$ . Practical profiles such as Gaussian-shaped profiles and a simplified superjunction design where the phosphorous implants were merged into the background SIC doping are proposed and examined. This novel technique might potentially be useful for all high-speed bipolar transistors.

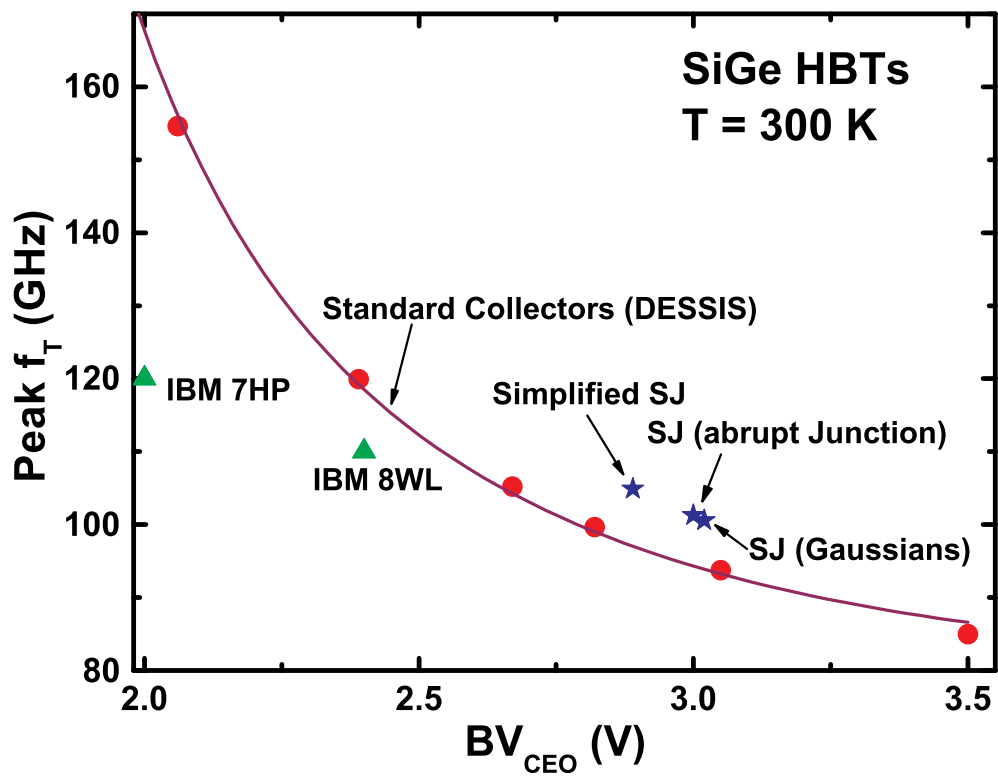


Figure 87: Benchmarking: peak  $f_T$  vs.  $BV_{CEO}$  for SiGe HBTs.

## CHAPTER VI

### MONTE CARLO SIMULATION

In this chapter, we explore full-band Monte Carlo (MC) simulation of silicon-based bipolar transistors using commercial Sentaurus SPARTA [180]. With technology scaling, electron transport in the CB SCR and the ultra-thin base region becomes quasi-ballistic, and a strong velocity overshoot occurs [186]. TCAD tools based on drift-diffusion (DD) cannot simulate velocity overshoot because DD assumes that electrons reach thermal equilibrium instantly with the lattice through scattering, while hydrodynamic models (HD) capture velocity overshoot under the assumption that the electron gas is slowly drifting and hot. However, such an assumption is questionable for ballistic transport where carriers become rapid drifting and cool. Therefore, more accurate solutions of the Boltzmann transport equation (BTE) become necessary in verifying the accuracy of HD or DD in each technology node of SiGe HBTs.

#### ***6.1 Issues with TCAD Tools***

Theoretically speaking, HD should be more accurate than DD due to the higher moments of BTE used. Unfortunately, HD overestimates the velocity overshoot and hence  $f_T$ , as compared to the direct solution of BTE [187], [188]. The extent of the overestimation varies from simulator to simulator. For example, [188] finds considerable overestimation of velocity overshoot in 75-GHz SiGe HBTs using and HD simulator, and the result from DD simulation [188] agrees well with MC simulation. Such a conclusion is against intuition, and is untrue in more modern TCAD tools (e.g. [182]). In our experience with Sentaurus Device (previously Dessis), HD is more accurate than DD in the simulation of second- and third-generation SiGe HBTs. Therefore, HD is used by most companies and research teams nowadays to assist the

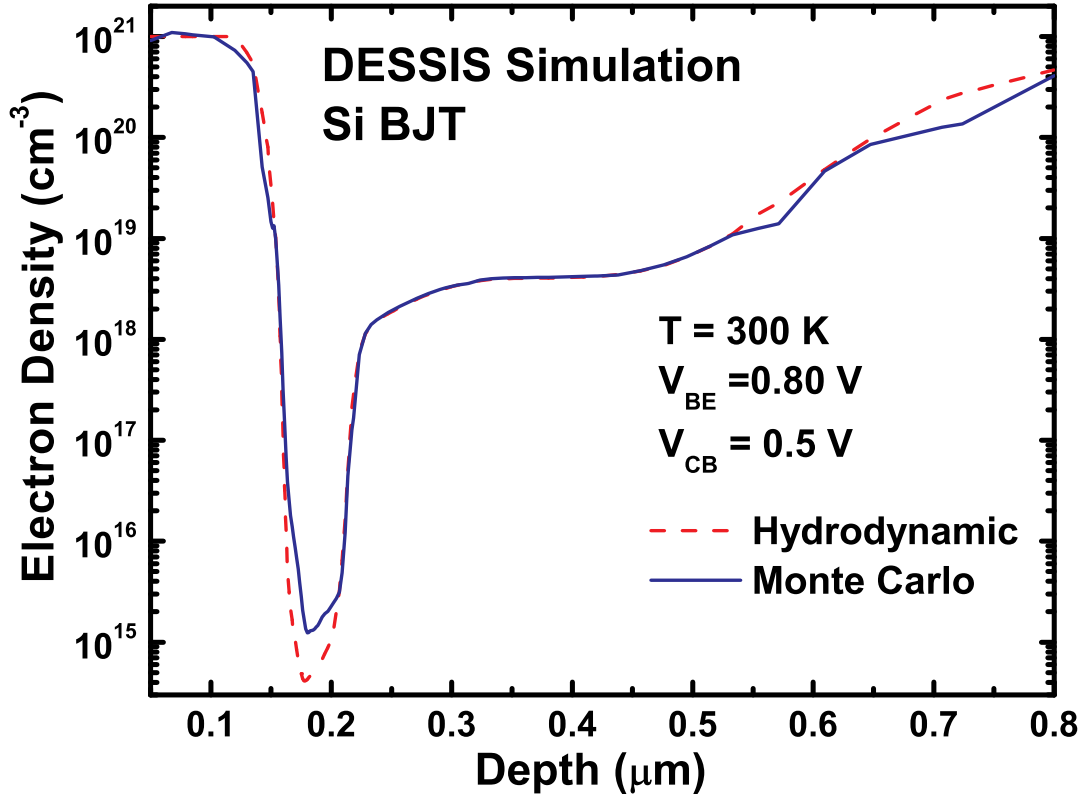
design of modern SiGe HBTs.

MC simulation of semiconductor devices has more than 40 years of history, and *Kurosawa* first applied MC techniques to high-field semiconductor transport in 1966 [179]. Since then, MC has been regarded as the most accurate way of solving the BTE, and tremendous progress has been made by numerous university groups and research labs, as summarized in [131]. In addition, commercial MC tools such as Sentaurus SPARTA and Sentaurus MOCA as described in [180] became available to commercial customers outside the small MC research community.

However, fully functional MC simulations are still unavailable in commercial MC tools for the following reasons. On the one hand, technical support of commercial MC tools is very costly because most users do not have the expertise to deal with the more complicated MC simulator as they normally do with HD/DD tools. On the other hand, companies are reluctant to pay for the technical support of MC simulators because: (1) Empirically tuning of HD parameters seems to be sufficient in regular technology development, and (2) the large numerical computation burden of MC simulation increases the production cycle. As a result, MC simulation is excluded from regular device engineering, despite the limitation of HD simulators. Consequently, even some functions that were claimed to be available in the Sentaurus MC manual are not truly functional. We will revisit this issue in the study of impact ionization.

It is, therefore, important to identify the technology node at which HD simulators are no longer adequate for device engineering with the aid of MC simulations. As mentioned in the beginning of this chapter, scaling of bipolar transistors challenges hydrodynamic models in at least two ways: velocity overshoot and breakdown behavior in the CB junction, both of which have been studied by MC simulations in [91], [181], and [182]. These papers, however, only studied devices slower than 100 GHz, and little progress has been made in the MC simulation of more advanced SiGe HBTs

where HD models become more questionable.



**Figure 88:** Simulated electron density *vs.* depth in a Si BJT with doping profiles equivalent to a 200-GHz SiGe HBT. The dashed line denotes hydrodynamic simulation, and the solid line shows Monte Carlo simulation.

It is also important to point out that the calibration of HD models that had been “proven” correct by existing technologies is quite tricky. What is routinely done in device engineering is to start with a SIMS profile and calibrate the profiles to match the measured data for each technology. Sometimes model parameters have to be tuned to match multiple curves (e.g.  $f_T$  *vs.*  $I_C$  and  $(M - 1)$  *vs.*  $V_{CB}$  curves), and those parameters are especially empirical when it comes to base current calibration. One can change carrier life time, tune surface recombination velocity, or manually create recombination/generation centers to reach the same base current. Similar flexibility

also exists in tuning the  $M - 1$  behavior. Iterations have to be made during the calibration process, and the final parameter set will be carried over as the starting point for the next generation technology. This is a self-consistent methodology; however, the error of the SIMS data themselves can be as high as 20%, and it can be even worse in doping retrogrades [39]. Therefore, the calibrated profile and the parameter sets may not be real. For instance, if the collector doping in HD simulation is lower than the actual doping, one may end up tuning the impact ionization parameters instead to match the  $M - 1$  performance.

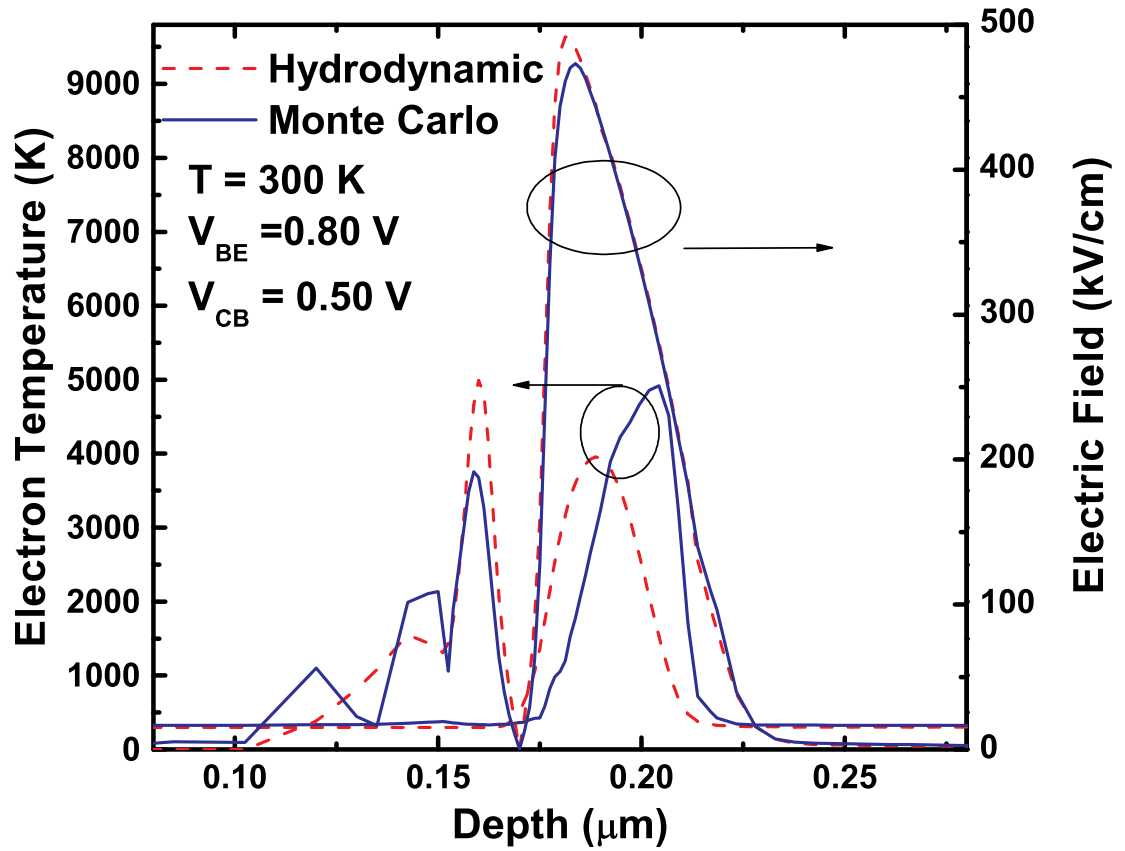
## 6.2 *Velocity Overshoot*

The same doping profile of a third-generation SiGe HBT is used in both Sentaurus MC and HD simulations for comparison. Germanium is not included in the simulations. Hence, we study Si BJTs to avoid the computation of SiGe composite of different germanium mole fractions. The simulation of a full SiGe HBT should involve making artificial staircases of SiGe with different germanium mole fractions whose band structures can be calculated by the Sentaurus band structure tool [180], and it is worth exploring in future research.

Convergence of MC simulation is achieved by setting MC time longer than 10 ms with 100 intervals (100  $\mu s$  in each interval) for Sentaurus SPARTA. The good agreement between HD and MC curves regarding electron density in Fig. 88 and electric field in Fig. 89 confirms that the simulation has converged.

In Fig. 90, velocity overshoot can be demonstrated by the peak values of the simulated electron velocity (about  $2.5 \times 10^7 \text{ cm} \cdot \text{s}^{-1}$ ) that is 2x higher than the thermal velocity ( $v_{th}$ ) and saturation velocity ( $v_{sat}$ ), both of which are about  $10^7 \text{ cm} \cdot \text{s}^{-1}$  in silicon. Note the physical meanings of  $v_{th}$  and  $v_{sat}$  are different as indicated in [93]. The extremely high electric field over the thin CB SCR is responsible for the velocity overshoot observed. HD and MC show good agreement in terms of the electron

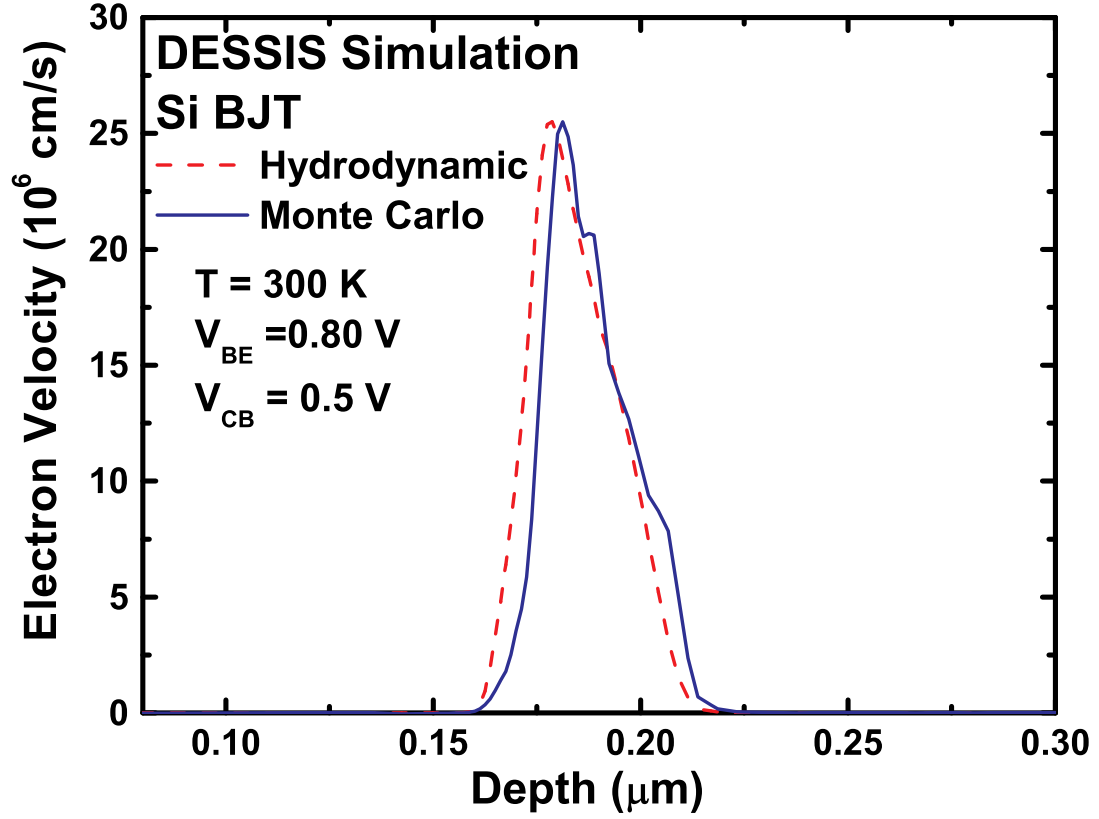




**Figure 89:** Simulated electron temperature (or electron energy) and electric field *vs.* depth in a Si BJT with doping profiles equivalent to a 200-GHz SiGe HBT. The dashed lines denote hydrodynamic simulation, and the solid lines show Monte Carlo simulation.

velocity profile, demonstrating that the HD simulation of device speed is still valid in the third-generation SiGe HBT technology node. According to our knowledge, this is the first time that the HD model has been verified by MC simulation in 200-GHz devices, and this conclusion is contrary to the previous benchmarking work in [188].

The dead-space effect can be observed in Fig. 89 where the location of the peak electron temperature (or electron energy) differs from that of the peak electric field. The electron temperature is an indication of electron kinetic energy, and higher kinetic energy causes stronger impact ionization. Although HD captures the dead-space effect, the length of the dead space is not consistent with the MC result.



**Figure 90:** Simulated electron velocity *vs.* depth in a Si BJT with doping profiles equivalent to a 200-GHz SiGe HBT. The dashed line denotes hydrodynamic simulation, and the solid line shows Monte Carlo simulation.

Let us further discuss the implication of Fig. 89. In terms of the HD simulation result, electron temperature reaches its maximum in the middle of the CB space charge region. This phenomena results from the balance between two forces: the accelerating electric field and the scattering inside CB SCR; an electron gains its energy as it travels through the electric field while losing energy through scattering. However, MC simulation clearly shows that the electron mainly gains energy throughout the entire space charge region, and then quickly loses energy through scattering at the edge of the space charge region. The different conclusions based on the two simulation methods are important for collector engineering in SiGe HBTs faster than 200 GHz.

### 6.3 Impact Ionization

Numerous assumptions in hydrodynamic models are questionable in more advanced SiGe HBTs, and the modeling of impact ionization is particular difficult transistors. The current model parameters used for HD impact ionization were calibrated in bulk silicon [177], under the assumption that electron temperature ( $T_n$ ) is the driving force of impact ionization. Ionization coefficient ( $\alpha_n$ ) is first calculated in bulk silicon undergoing certain electric field. The electric field is then converted to  $E_{eff}$  that is a function of  $T_n$  of the bulk silicon. Consequently, the empirical  $\alpha_n$  *vs.*  $T_n$  can be determined as shown in Equation (27) in Chapter 5. Such an assumption, however, is too bold, because transport in bulk silicon is significantly different than that in thin silicon layers. Spatial velocity overshoot that happens in the CB junction cannot be fully understood if one only looks into bulk silicon. Moreover,  $T_n$  is a convenient approximation only. It represents the electron kinetic energy, and is less meaningful under high field, when electron distribution becomes severely non-Maxwellian [130]. Consequently, MC simulation of impact ionization is more reliable than HD.

#### 6.3.1 Limitations of Hydrodynamic Models

To understand the complexity of the problem mentioned above, let us first revisit the concept of electron temperature in the derivation of hydrodynamic models by following the analyses in [183] and [184]. By substituting Eq. (4) into Eq. (6) in Chapter 1 and expanding the energy flux term  $J_W$ , one reaches Equation (29),

$$n \frac{\partial w}{\partial t} + \nabla_r \cdot \vec{Q} + n \vec{v} \cdot \nabla_r w + \nabla_r \cdot (nk\hat{T}\vec{v}) + qn\vec{v} \cdot \vec{E} = C_W - (G_n - R_n)w, \quad (29)$$

where  $w$  is the average energy,  $\vec{v}$  the average velocity,  $\vec{Q}$  the heat flux, and  $\hat{T}$  the temperature tensor. The electron velocity consists of two components, or,  $\vec{u} = \vec{c} + \vec{v}$ ,

in which  $\vec{c}$  is the random component of the electron velocity. Therefore, the heat flux can be expressed as:

$$\vec{Q} = \int \frac{1}{2} m c^2 \vec{c} f d^3 u. \quad (30)$$

Each element of the temperature tensor is defined as:

$$n k T_{ij} = \int m c_i c_j f d^3 u. \quad (31)$$

The average electron energy can then be found as:

$$w = \frac{\int \frac{1}{2} m u^2 f d^3 u}{\int f d^3 u} = \frac{1}{2} m v^2 + \frac{1}{2} k (T_{11} + T_{22} + T_{33}). \quad (32)$$

Three assumptions have to be made to simplify Eq. (29).

Assumption #1: the tensor  $\hat{T}$  collapses to a scalar quantity, or  $T_{11} = T_{22} = T_{33} = T_n$ , and this is the definition of the electron temperature that we used in the hydrodynamic simulation.

Assumption #2: the relaxation time approximation is a constant, that is,

$$C_W - (G_n - R_n) w = -n \frac{w - w_0}{\tau_W}, \quad (33)$$

where  $w_0 = \frac{3}{2} k T_L$ , and  $T_L$  is the lattice temperature.

Assumption #3:  $\vec{Q} = -k_n \nabla_r \cdot T_n$ , where  $k_n$  is the thermal conductivity.

Under these three assumptions, Eq. (29) can be simplified as:

$$\nabla_r \cdot \left( k_n \nabla T_n + \frac{1}{q} (w + k T_n) \vec{J}_n \right) + \vec{E} \cdot \vec{J}_n + (G_n - R_n) w = n \frac{w - w_0}{\tau_W}. \quad (34)$$

Eq. (34) is the fundamental equation used in most of the hydrodynamic/energy balance simulators, where  $T_n$  is also used as the driving force for all impact ionization models with hydrodynamics [177]. Unfortunately, however, none of the three assumptions mentioned above are justified under high field: The tensor  $\hat{T}$  cannot be simplified as a scalar, as electrons have significantly higher energies along the direction of the electric field; the assumption that  $k_n$  and  $\tau_W$  are constant is for electrons with high kinetic energies.

Most importantly, the temperature tensor  $\hat{T}$  itself in Eq. 32 loses key information under high field: the form (shape) of the distribution function. The connotation of using such a parameter is that carrier distribution is approximated to be Maxwellian (or more strictly speaking, the displaced Maxwellian distribution), and the distribution function,  $f$ , can be fully expressed by one parameter:  $T$ . However, as the distribution function becomes highly asymmetric and distorted (or non-Maxwellian) in high-field transport,  $\hat{T}$  alone conveys no information beyond the total kinetic energy, and the distribution function is not defined. Understanding of this matter provides insight into why impact ionization parameters extracted from bulk silicon cannot be used in highly-scaled semiconductor transistors where abrupt spatial changes of electric field produces different distribution functions that do not apply bulk silicon.

The current (temporary) solution in HD simulations is to calibrate those parameters in hydrodynamic models for each technology node. Such a temporary parameter set, nevertheless, cannot be predictive for future technologies. Therefore, Monte Carlo simulation seems to be the best and the only solution for studying impact ionization in device scaling.

### 6.3.2 Sentaurus Monte Carlo Simulation

Sentaurus Device has two Monte Carlo simulators: MOCA and SPARTA. The former does ensemble Monte Carlo (EMC) simulation while the latter simulates one

particle after another. The difference between EMC and single-particle MC is important in non-ergodic systems [131]. In MOCA, one defines the simulated carrier type by setting “carrier=e” or “h” in the “math” section; whereas SPARTA integrates pre-simulated density distributions over the entire Monte Carlo window and simulates only the carrier type with the greater integral of the density. Either way, electrons are simulated in NPN transistors.

According to the user’s manual [180], one can set the impact ionization model in MOCA by using “impact\_ionization\_model=Cartier” in the “physics” section, and then set all the parameters in the “impact” section. One can also choose Bude’s model or Kane’s model as the impact ionization model. SPARTA handles avalanche with less complexity as well as less flexibility; only Catier’s model is available, but one only needs to set “IIFactor=1.0” to switch on impact ionization.

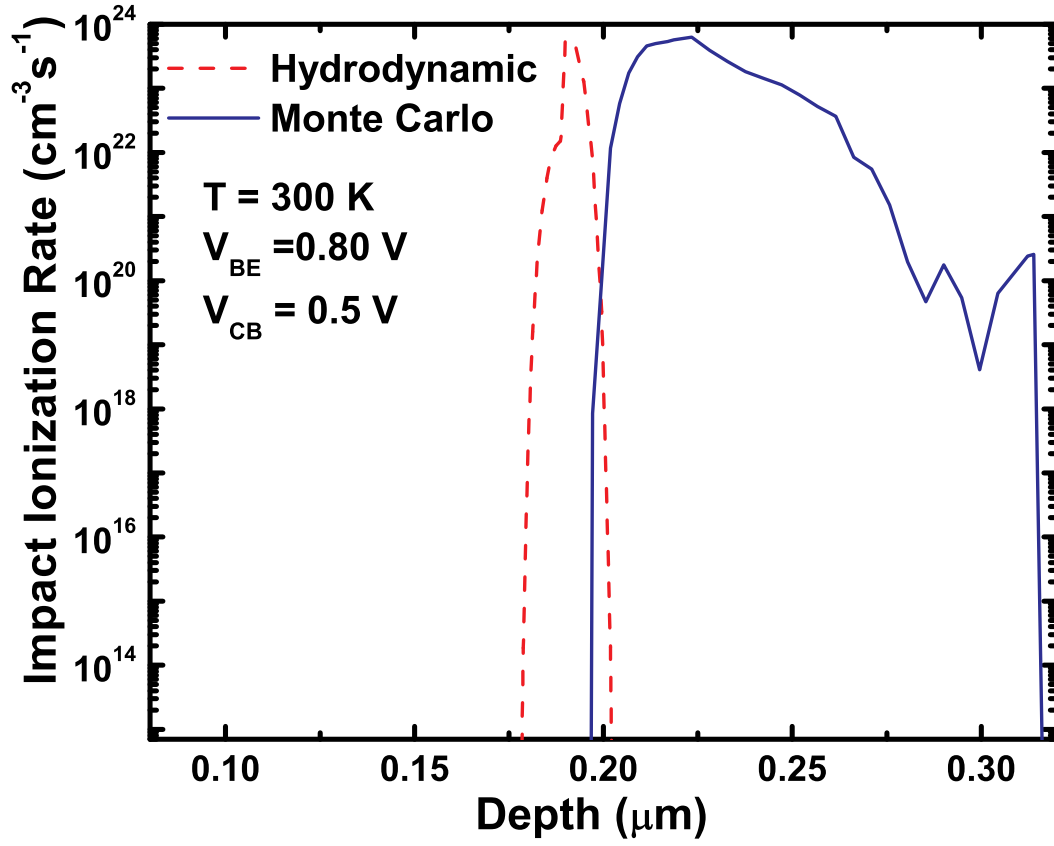
Ideally, after impact ionization is simulated, one can derive the device multiplication factor through its impact ionization generation rate ( $G_n^{ii}$ ), e.g. the parameter “eMCAvalanche” in SPARTA. The ionization coefficient can be calculated using:

$$G_n^{ii} = nv\alpha_n. \quad (35)$$

Then  $M - 1$  across the CB junction can be extracted using the conclusion in [184]:

$$M - 1 = e^{\int_0^W \alpha_n dx} - 1. \quad (36)$$

Unfortunately, neither of the current MC simulators in Sentaurus is able to simulate impact ionization, despite what is claimed in the manuals. Impact ionization is a multi-particle process, and it cannot be simulated if only one kind of particle is simulated. As Professor Yoder indicated, one could in principle implement 2-particle simulation and then iterate, but this function is unavailable at the moment. This conclusion can also be verified by comparing the impact ionization rate simulated in the HD simulator with that in the MC simulator (Fig. 91). In this figure, HD makes



**Figure 91:** Simulated electron impact ionization rate *vs.* depth in a Si BJT with doping profiles equivalent to a 200-GHz SiGe HBT. The dashed line denotes hydrodynamic simulation, and the solid line shows Monte Carlo simulation.

more sense than MC, because strong impact ionization is expected to take place in CB SCR. It is strongly recommended that future commercial MC tools should include impact ionization.

### 6.3.3 Zener Tunneling

Another important mechanism for breakdown is Zener tunneling. Device engineers have been paying attention to impact ionization while ignoring the Zener process in the study of the CB junction, mainly because the collector doping has not been high enough to cause significant Zener tunneling. With collector doping above  $10^{18} \text{ cm}^{-3}$  in SiGe HBTs near half-THz performance, Zener tunneling becomes significant [185].

There are evidence of the Zener process even in slower SiGe HBTs. In the work reported in [181], the authors have to manually increase the phonon scattering rates in the MC simulator by 6% to match measured  $M - 1$  data. It is likely that this extra current component is due to Zener tunneling. A more direct evidence can be shown in our cryogenic measurement of fourth-generation SiGe HBTs that are reported in Chapter 4. The  $BV_{CBO}$  of these transistors remains at 5.6 V at a temperature range of 4.5 K to 300 K. Traditionally,  $BV_{CBO}$  is expected to have a positive temperature coefficient if the breakdown behavior is caused by avalanching (because impact ionization is less effective when the mean free path is shorter at higher temperature), and a slight positive temperature coefficient if breakdown is dominated by the Zener process. Therefore, the measured constant  $BV_{CBO}$  against temperature results from a combination of both Zener and avalanching.

Zener tunneling is expected to play a more important role with scaling. Meanwhile, current device simulators (HD or MC) cannot capture Zener process, and this function is highly desirable in future TCAD study for SiGe HBTs.

## 6.4 *Summary*

This chapter studies velocity overshoot and impact ionization by comparing MC and HD simulation results. It is demonstrated that the HD models used in Sentaurus Device Simulator gives decent results in terms of velocity overshoot in a third-generation SiGe HBT; however, the dead-space effect is more significant in the MC simulation, revealing a different acceleration-scattering balancing process. Roadblocks of HD and MC simulations of THz SiGe HBTs have been identified, and the ability to simulate impact ionization and Zener tunneling is highly desirable in future commercial MC tools.



## CHAPTER VII

### CONCLUSION AND FUTURE WORK

This thesis has studied the behavior of SiGe HBTs at cryogenic temperatures and its relation to device scaling and optimization. One can explain a novel negative differential resistance (NDR) effect and a collector current kink effect in the first-generation SiGe HBTs at deep cryogenic temperature within the traditional drift-diffusion framework. A theory of positive feedback due to the enhanced heterojunction barrier effect at deep cryogenic temperature is proposed to explain such effects. This theory instructs the profile optimization of the germanium and base doping profiles to both suppress carrier freezeout and the heterojunction barrier effect at deep cryogenic temperatures, leading to a significant improvement in the DC and RF performance for NASA lunar missions.

Cooling is further used as a tuning knob to better understand the performance limits of SiGe HBTs. The consequences of cooling SiGe HBTs are in many ways similar to those of combined vertical and lateral device scaling. A case study of low-temperature DC and RF performance of prototype fourth-generation SiGe HBTs is presented. It summarizes the performance of all three prototypes of these fourth-generation SiGe HBTs within the temperature range of 4.5 to 300 K. This work helps to analyze the key optimization issues associated with device scaling to terahertz speeds at room-temperature. Record performance is also achieved in a CML ring oscillator with 2.3 ps gate delay operating at cryogenic temperatures. Decreased intrinsic base sheet resistance is measured and contributes to the high  $f_{max}$  and the short gate delay.

As an alternative method, an  $f_T$ -doubler technique is presented as an attempt

to reach half-terahertz speeds. In addition, a roadmap for terahertz device scaling is given, and the potential relevant physics that will necessarily be encountered in future device scaling are examined, including higher base recombination current component and non-equilibrium base transport. TCAD simulation using hydrodynamic models confirms the possibility of reaching THz performance in SiGe HBTs. Subsequently, a novel superjunction collector design is proposed for higher breakdown voltages. Such a novel method was used in 100-GHz SiGe HBTs to improve breakdown voltages. Practical profiles were designed for actual device engineering. At last, Monte Carlo simulations are explored to study velocity overshoot and impact ionization of aggressively-scaled SiGe HBTs.

Carrier transport in highly-scaled or cooled SiGe HBTs is an open topic, and the direct solution of the Boltzmann transport equation is becoming increasingly important for these transistors. For future directions, SiGe HBTs will continue to play a major role in extreme environment applications (high radiation and/or low (or high) temperatures). Emerging markets for SiGe HBTs include all high-speed IC applications and possibly low-power electronics.

## REFERENCES

- [1] D. L. Harame, D. C. Ahlgren, D. D. Coolbaugh, J. S. Dunn, G. Freeman, J. D. Gillis, R. A. Groves, G. N. Hendersen, R. A. Johnson, A. J. Joseph, S. Subbanna, A. M. Victor, K.M. Watson, C. S. Webster, and P. J. Zampardi, "Current status and future trends of SiGe BiCMOS technology," *IEEE Transactions on Electron Devices*, vol. 48, pp. 2575-2594, 2001.
- [2] J. D. Cressler, "On the potential of SiGe HBTs for extreme environment electronics," *Proceedings of the IEEE*, vol. 93, pp. 1559-1582, 2005.
- [3] J. D. Cressler, R. Krithivasan, G. Zhang, G. Niu, P. Marshall, H. Kim, R. Reed, M. Palmer, and A. Joseph, "An investigation of the origins of the variable proton tolerance in multiple SiGe HBT BiCMOS technology generations," *IEEE Transactions on Nuclear Science*, vol. 49, pp. 3203-3207, 2002.
- [4] Y. Lu, J. D. Cressler, R. Krithivasan, Y. Li, R. A. Reed, P.W. Marshall, C. Polar, G. Freeman, and D. Ahlgren, "Proton tolerance of third-generation  $0.12\mu\text{m}$  185 GHz SiGe HBTs," *IEEE Transactions on Nuclear Science*, vol. 50, pp. 1811-1815, 2003.
- [5] J. D. Cressler, J. H. Comfort, E. F. Crabbe, G. L. Patton, J. M. C. Stork, J. Y. -C. Sun, and B. S. Meyerson, "On the profile design and optimization of epitaxial Si- and SiGe-base bipolar technology for 77 K applications. I. transistor DC design considerations," *IEEE Transactions on Electron Devices*, vol. 40, pp. 525-541, 1993.
- [6] J.D. Cressler, E.F. Crabbe, J.H. Comfort, J.M.C. Stork, J.M.C. and J.Y.-C. Sun, "On the profile design and optimization of epitaxial Si- and SiGe-base bipolar technology for 77 K applicationpart II: circuit performance issues," *IEEE Transactions on Electron Devices*, vol. 40, pp. 542-556, 1993.
- [7] B. Banerjee, S. Venkataraman, Y. Yuan, Q. Liang, C.-H. Lee, S. Nuttinck, D. Heo, Y.-J. Chen, J. D. Cressler, J. Laskar, G. Freeman, and D. C. Ahlgren, "Cryogenic operation of third-generation, 200-GHz peak- $f_T$ , silicon-germanium heterojunction bipolar transistors," *IEEE Transactions on Electron Devices*, vol. 52, pp. 585-593, 2005.
- [8] D. L. Harame, J. H. Comfort, J. D. Cressler, E. F. Crabbe, J. Y. -C. Sun, B. S. Meyerson, and T. Tice, "Si/SiGe epitaxial-base transistors. I. materials, physics, and circuits," *IEEE Transactions on Electron Devices*, vol. 42, pp. 455-468, 1995.
- [9] A. J. Joseph, J. D. Cressler, and D. M. Richey, "Operation of SiGe heterojunction bipolar transistors in the liquid-helium temperature regime," *IEEE Electron Device Letters*, vol. 16, pp. 268-170, 1995.

- [10] D. C. Ahlgren, M. Gilbert, D. Greenberg, S. J. Jeng, J. Malinowski, D. Nguyen-Ngoc, K. Schonenberg, K. Stein, R. Groves, K. Walter, G. Hueckel, D. Colavito, G. Freeman, D. Sunderland, D. L. Haramé, and B. Beyerson, "Manufacturability demonstration of an integrated SiGe HBT technology for the analog and wireless marketplace," in *Tech. Dig. International Electron Devices Meeting*, pp. 859-862, 1996.
- [11] A. Joseph, D. Coolbaugh *et al.*, "A 0.18 $\mu$ m BiCMOS technology featuring 120/100 GHz ( $f_T$  /  $f_{max}$ ) HBT and ASIC-compatible CMOS using copper interconnect," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 143-146, 2001.
- [12] B. Jagannathan, M. Khater, F. Pagette, J. -S. Rieh, D. Angell, H. Chen, J. Florkey, F. Golan, D. R. Greenberg, R. Groves, S. J. Jeng, J. Johnson, E. Mengitsu, K. T. Schonenberg, C. M. Schnabel, P. Smith, A. Stricker, D. Ahlgren, G. Freeman, K. Stein, and S. Subanna, "Self-aligned SiGe NPN transistors with 285 GHz  $f_{max}$  and 207 GHz  $f_T$  in a manufacturable technology," *IEEE Electron Device Letters*, vol. 23, pp. 258-260, 2002.
- [13] J.-S. Rieh, B. Jagannathan *et al.*, "SiGe HBTs with cut-off frequency of 350 GHz," in *Tech. Dig. IEEE International Electron Devices Meeting*, pp. 771-774, 2002.
- [14] J.-S. Rieh, D. Greenberg *et al.*, "SiGe HBTs for millimeter-wave applications with simultaneously optimized  $f_T$  and  $f_{max}$  of 300 GHz," in *Proc. IEEE Radio Frequency Integrated Circuits Symposium*, pp. 395-398, 2004.
- [15] M. Khater, J.-S. Rieh *et al.*, "SiGe HBT technology with  $f_{max}/f_T = 350/300$  GHz and gate delay below 3.3 ps," in *Tech. Dig. IEEE International Electron Devices Meeting*, pp. 247-250, 2004.
- [16] P. Chevalier, F. Pourchon, T. Lacave, G. Avenier, Y. Campidelli, L. Depoyan, G. Troillard, M. Buczko, D. Gloria, D. Cli, C. Gaquire, and A. Chantre, "A Conventional double-polysilicon FSA-SEG Si/SiGe:C HBT reaching 400 GHz  $f_{MAX}$ ," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, paper 1.1, Oct. 2009.
- [17] S. van Huylenbroeck, A. Sibaja-Hernandez, R. Venegas, S. You, G. Winderickx, D. Radisic, W. Lee, P. Ong, T. Vandeweyer, D. Nguyen, K. De Meyer, and S. Decoutere, "A 400 GHz  $f_{MAX}$  fully self-aligned SiGe:C HBT architecture," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, paper 1.2, Oct. 2009.
- [18] J. Yuan, C. Zhu, Y. Cui, J.D. Cressler, G. Niu, Q. Liang, E. Zhao, A. Appaswamy, R. Krithivasan, and A. Joseph, "A New device phenomenon in cryogenically-operated SiGe HBTs," in *Tech. Dig. IEEE Int. Electron Devices Meeting*, pp. 611-614, Dec. 2006.

- [19] J. Yuan, J. D. Cressler, C. Zhu, Y. Cui, G. Niu, Q. Liang, and A. J. Joseph, "An investigation of negative differential resistance and novel collector current kink effects in SiGe HBTs operating at cryogenic temperatures," *IEEE Transactions on Electron Devices*, Vol. 54, No. 3, pp. 504-516, Mar. 2007.
- [20] A. Appaswamy, M. Bellini, W.-M. L. Kuo, P. Cheng, J. Yuan, C. Zhu, J. D. Cressler, G. Niu, A. J. Joseph, "Impact of scaling on the inverse-mode operation of SiGe HBTs," *IEEE Transactions on Electron Devices*, Vol. 54, No. 6, pp. 1492-1501, June 2007.
- [21] J. Yuan, J. D. Cressler *et. al.*, "On the performance limits of cryogenically-operated SiGe HBTs and its relation to scaling for terahertz speeds," *IEEE Transactions on Electron Devices*, vol. 56, no. 5, pp. 1007-1019, May 2008.
- [22] J. Yuan, K.A. Moen *et. al.*, "SiGe HBT CML ring oscillator with 2.3 ps gate delay at cryogenic temperatures," *IEEE Transactions on Electron Devices*, in press.
- [23] J. Yuan, R. Krithivasan *et al.*, "On the frequency limits of SiGe HBTs for Terahertz applications," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 22-25, Oct. 2007.
- [24] S. Finn, J. Yuan *et al.*, "A 10 Mbps SiGe BiCMOS transceiver for operation down to cryogenic temperatures," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 115-118, Oct. 2007.
- [25] J. Yuan, J. D. Cressler *et al.*, "On the profile design of SiGe HBTs for RF lunar applications down to 43 K," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 25-28, Oct. 2008.
- [26] L. Luo, G. Niu, D. Thomas, J. Yuan, and J. D. Cressler, "Forced- $I_E$  pinch-in maximum output voltage limit in SiGe HBTs operating at cryogenic temperatures," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 41-44, Oct. 2008.
- [27] J. Yuan and John D. Cressler, "A novel superjunction collector design for improving breakdown voltage in high-speed SiGe HBTs," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, Capri, Italy, paper 5.2, Oct. 2009.
- [28] J. Yuan and J. D. Cressler, "Enhancing the speed of SiGe HBTs Using  $f_T$ -Doubler Techniques," in *Proc. IEEE Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems*, pp. 50-53, Feb. 2008.
- [29] T. K. Thrivikraman, J. Yuan *et. al.*, "SiGe HBT X-band LNAs for ultra-low-noise cryogenic receivers," *IEEE Microwave Theory and Wireless Component Letters*, vol. 18, no. 7, pp. 476-478, 2008.
- [30] Jiahui Yuan, "SiGe HBTs Operating at Deep Cryogenic Temperatures," Master Thesis, Georgia Institute of Technology, Atlanta, GA, 2007.

- [31] W. Shockley, "Semiconductor Amplifier," US Patent 2,502,488, issued Apr. 1950.
- [32] H. Kroemer, "Theory of a wide-gap emitter for transistors," in *Proc. IRE*, vol. 45, no. 11, pp. 1535-1537, Nov. 1957.
- [33] G. O. Ladd and D. L. Feucht, "Performance potential of high-frequency heterojunction transistors," *IEEE Transactions on Electron Devices*, vol. 17, pp. 413-420, May 1970.
- [34] H. Kroemer, "Heterostructure bipolar transistors and integrated circuits," in *Proc. of the IEEE*, vol. 70, No. 1, pp. 13-25, Jan. 1982.
- [35] D. Ankri and A. Scavennec, "Design and evaluation of a planar GaAlAs-GaAs bipolar transistor," *Electronic Letters*, vol. 16, no. 1, pp. 41-47, Jan. 1980.
- [36] J.-P. Bailbe, A. Marty, P. H. Hiep, and G. E. Rey, "Design and fabrication of high-speed GaAlAs/GaAs heterojunction transistors," *IEEE Transactions on Electron Devices*, vol. ED-27, pp. 1160-1164, June 1980.
- [37] M. Tobe, Y. Amemiya, S. Sakai, and M. Umeno, "High-sensitivity InGaAsP/InP phototransistors," *Applied Physics Letters*, vol. 37, no. 1, pp. 73-75, July 1980.
- [38] Y. Taur, T. H. Ning, *Fundamentals of Modern VLSI Devices*. Cambridge, U.K.: Cambridge University Press, 1998.
- [39] J. D. Cressler and G. Niu, *Silicon-Germanium Heterojunction Bipolar Transistors*. Boston, MA: Artech House, 2003.
- [40] J. S. Rieh, B. Jagannathan, D. R. Greenberg, M. Meghelli, A. Rylyakov, F. Guarin, Z. Yang, D. C. Ahlgren, G. Freeman, P. Cottrell, and D. Hareme, "SiGe heterojunction bipolar transistors and circuits toward terahertz communication applications," *IEEE Trans. Microwave Theory and Techniques*, vol. 52, pp. 2390-2407, 2004.
- [41] P. Chevalier, N. Zerounian, B. Barbalat, F. Aniel and A. Chantre, "On the use of cryogenic measurements to investigate the potential of Si/SiGe:C HBTs for terahertz application," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 26-29, 2007.
- [42] N. Zerounian, E. Ramirez Garcia, F. Aniel, P. Chevalier, B. Geynet, and A. Chantre, "SiGe HBT featuring  $f_T$  600 GHz at cryogenic temperature," *ECS Transactions*, vol. 16, no. 10, pp. 1069-1077, 2008.
- [43] R. Krithivasan, Y. Lu, J. D. Cressler, J. -S. Rieh, M. H. Khater, D. Ahlgren, and G. Freeman, "Half-terahertz operation of SiGe HBTs," *IEEE Electron Device Letters*, vol. 27, pp. 567-569, 2006.

- [44] Q. Liang, R. Krithivasan, A. Ahmed, Y. Lu, Y. Li, J. D. Cressler, G. Niu, J. -S. Rieh, G. Freeman, D. Ahlgren, and A. Joseph, "Analysis and understanding of unique cryogenic phenomena in state-of-the-art SiGe HBTs," *Solid State Electronics*, vol. 50, pp. 964-972, 2006.
- [45] W. Webster, "On the variation of junction-transistor current-amplification factor with emitter current," *Proceedings of the IRE*, vol. 42, no. 6, pp. 914-920, June 1954.
- [46] E. S. Rittner, "Extension of the theory of the junction transistor," *Physical Review*, vol. 94, pp. 1161-1171, 1954.
- [47] J. Kirk, C.T., "A theory of transistor cutoff frequency  $f_T$  falloff at high current densities," *IRE Transactions on Electron Devices*, vol. 9, no. 2, pp. 164-174, March 1962.
- [48] J.C. Bardin, "Silicon-germanium heterojunction bipolar transistors for extremely low-noise applications," Ph.D. dissertation, California Institute of Technology, Pasadena, CA, 2009.
- [49] S. Tiwari, "A new effect at high currents in heterostructure bipolar transistors," *IEEE Electron Device Letters*, vol. 9, pp. 142-144, 1988.
- [50] S. Tiwari and D. J. Frank, "Analysis of the operation of GaAlAs/GaAs HBTs," *IEEE Transactions on Electron Devices*, vol. 36, pp. 2105-2121, 1989.
- [51] Z. Yu, P. E. Cottrell, and R. W. Dutton, "Modeling and simulation of high-level injection behavior in double heterojunction transistors," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 192-194, 1990.
- [52] J. D. Cressler, D. M. Richey, R. C. Jaeger, E. F. Crabe, J. H. Comfort, and J. M. C. Stock, "High-injection barrier effects in SiGe HBTs operating at cryogenic temperatures," *Journal De Physics IV*, vol. 4, pp. C6-117-122, 1994.
- [53] A. J. Joseph, J. D. Cressler, D. M. Richey, and G. Niu, "Optimization of SiGe HBTs for operation at high current densities," *IEEE Transactions on Electron Devices*, vol. 46, pp. 1347-1354, 1999.
- [54] Q. Liang, J. D. Cressler, G. Niu, R. M. Malladi, K. Newton, and D. L. Hareme, "A physics-based high-injection transit-time model applied to barrier effects in SiGe HBTs," *IEEE Transactions on Electron Devices*, vol. 49, pp. 1807-1813, 2002.
- [55] P. H. Siegel, "Terahertz technology," *IEEE Transactions on Microwave Theory and Technology*, vol. 50, no. 3, pp. 910-928, March 2002.
- [56] M.J.W. Rodwell, M. Urteaga, T. Mathew *et al.*, "Submicron scaling of HBTs," *IEEE Transactions on Electron Devices*, vol. 48, no. 11, pp. 2606-2624, Nov. 2001.

- [57] J.C. Candy and G.C. Temes, *Oversampling Delta-Sigma Data Converters*, Piscataway, NJ: IEEE Press, 1992.
- [58] E. R. Mueller, "Terahertz radiation: Applications and sources," *The Industrial Physicist*, pp. 27-29, Aug./Sep. 2003.
- [59] P. D. Coleman and R. C. Becker, "Present state of the millimeter wave generation and technique art1958," *IEEE Transactions on Microwave Theory and Technology*, vol. 7, no. 1, pp. 42-61, Jan. 1959.
- [60] P. D. Coleman, "State of the art: Background and recent developments millimeter and submillimeter waves," *IEEE Transactions on Microwave Theory and Technology*, vol. MTT-11, no. 5, pp. 271-288, Sep. 1963.
- [61] J. C. Wiltse, "History of millimeter and submillimeter waves," *IEEE Transactions on Microwave Theory and Technology*, vol. MTT-32, no. 9, pp. 1118-1127, Sep. 1984.
- [62] D. M. Mittleman, M. Gupta, R. Neelamani, R. G. Baraniuk, J. V. Rudd, and M. Koch, "Recent advances in terahertz imaging," *Appl. Phys. B, Photophys. Laser Chem.*, vol. 68, no. 6, pp. 1085-1094, 1999.
- [63] J.-S. Rieh, D. Greenberg, A. Stricker, and G. Freeman, "Scaling of SiGe heterojunction bipolar transistors," *Proceedings of the IEEE*, vol. 93, no. 9, pp. 1522-1538, 2005.
- [64] S. Weinreb, J. C. Bardin, and H. Mani, "Design of cryogenic SiGe low noise amplifiers," *IEEE Trans. Microwave Theory and Techniques*, vol. 55, no. 11, pp. 2306-2312, 2007.
- [65] R. Lai, X. B. Mei, W. R. Deal, W. Yoshida, Y. M. Kim, P. H. Lin *et al.*, "Sub 50 nm InP HEMT device with  $f_{max}$  greater than 1 THz," in *Tech. Dig. IEEE International Electron Devices Meeting*, pp. 609-611, 2007.
- [66] S.-J. Yeon, M. Park, J. H. Choi, and K. Seo, "610 GHz InAlAs/In<sub>0.75</sub> GaAs metamorphic HEMTs with an ultra-short 15 nm gate," in *Tech. Dig. IEEE International Electron Devices Meeting*, pp. 613-616, 2007.
- [67] W. Snodgrass, W. Hafez, N. Harff, and M. Feng, "Pseudomorphic InP/InGaAs heterojunction bipolar transistors (PHBTs) experimentally demonstrating  $f_T = 765$  GHz at 25 C increasing to  $f_T = 845$  GHz at -55 C," in *Tech. Dig. of the IEEE International Electron Devices Meeting*, pp. 22.1, 2006.
- [68] Z. Griffith, E. Lind, and M. J. W. Rodwell, "Sub-300nm InGaAs/InP type-I DHBTs with a 150 nm collector, 30 nm base demonstrating 755 GHz  $f_{max}$  and 416 GHz  $f_T$ ," in *Proc. IEEE/LEOS International Conf. Indium Phosphide and Related Materials*, Japan, 2007.



- [69] B. Floyd, U. Pfeiffer, S. Reynolds, A. Valdes-Garcia, C. Haymes, Y. Katayamat, D. Nakanot, T. Beukema, B. Gaucher, and M. Soyuer, "Silicon Millimeter-Wave Radio Circuits at 60-100 GHz," in *Proc. Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems*, pp. 213-218, 2007.
- [70] A. Komijani and A. Hajimiri, "A wideband 77 GHz, 17.5 dBm power amplifier in silicon," in *Proc. IEEE Custom Integrated Circuits Conference*, pp. 566-569, Sept. 2005.
- [71] Laskin, M. Khanpour, R. Aroca, K. W. Tang, P. Garcia, and S. P. Voinigescu, "A 95 GHz receiver with fundamental frequency VCO and static frequency divider in 65nm digital CMOS," in *Tech. Dig. Solid-State Circuits Conference*, pp. 180-181, 2008.
- [72] B. Afshar, Y. Wang, and A. M. Niknejad, "A robust 24 mW 60 GHz receiver in 90nm standard CMOS," in *Tech. Dig. Solid-State Circuits Conference*, pp. 182-183, 2008.
- [73] S. Reynolds, B. Floyd, U. Pfeiffer, T. Beukema, J. Grzyb, C. Haymes, B. Gaucher, and M. Soyuer, "A silicon 60GHz receiver and transmitter chipset for broadband communications," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 12, pp. 2820-2830, Dec. 2006.
- [74] Y. Katayama, C. Haymes, D. Nakano, T. Beukema, B. Floyd, S. Reynolds, U. Pfeiffer, B. Gaucher, and K. Schleupen, "2-Gbps uncompressed HDTV transmission over 60-GHz SiGe radio link," in *Proc. IEEE Consumer Communications and Networking Conference*, pp. 12-16, Jan. 2007.
- [75] U. R. Pfeiffer and D. Goren, "A 20 dBm fully-integrated 60 GHz SiGe power amplifier with automatic level control," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 7, pp. 1455-1463, July 2007.
- [76] S. K. Reynolds, B. A. Floyd, U. R. Pfeiffer, T. J. Beukema, T. Zwick, J. Grzyb, D. Liu, and B. P. Gaucher, "Progress toward a low-cost millimeter-wave silicon radio," in *Proc. IEEE Custom Integrated Circuits Conference*, pp. 563- 570, Sept. 2005.
- [77] S. K. Reynolds and J. Powell, "77 and 94-GHz downconversion mixers in SiGe BiCMOS," in *Proc. IEEE Asian Solid-State Circuits Conference*, pp. 191-194, Nov. 2006.
- [78] B. Gaucher, B. Floyd, S. Reynolds, U. Pfeiffer, J. Grzyb, A. Joseph, E. Mina, B. Orner, H. Ding, R. Wachnik, and K. Walter, "Silicon germanium based millimeter-wave ICs for Gbps wireless communications and radar systems," *Semiconductor Science and Technology*, vol. 22, no. 1, pp. S236-S243, Jan. 2007.
- [79] I. Post, M. Akbar *et al.* "A 65 nm CMOS SOC technology featuring strained silicon transistors for RF applications," in *Tech. Dig. IEEE International Electron Devices Meeting*, paper 17.7, 2006.

- [80] A. Fox, B. Heinemann, R. Barth, D. Bolze, J. Drews, U. Haak, D. Knoll, B. Kuck, R. Kurps, S. Marschmeyer, H.H. Richter, H. Rucker, P. Schley, D. Schmidt, B. Tillack, G. Weidner, C. Wipf, D. Wolansky, Y. Yamamoto, "SiGe HBT module with 2.5ps gate delay," in *Tech. Dig. International Electron Devices Meeting*, pp. 731-734, Dec. 2008.
- [81] N.K. Srivastava, G. Raghavan, R. Thiagarajah, M.G. Case, E. Arnold, C.W. Pobanz, S.O. Nielsen, J.C. Yen, and R.A. Johnson, "InP HBT ring oscillator with 2.0 ps/stage gate delay," in *Tech. Dig. IEEE Gallium Arsenide Integrate Circuit Symposium*, pp. 171-174, 2002.
- [82] J.D. Cressler, E.F. Crabbe, J.H. Comfort, J.Y.-C. Sun, and J.M.C. Stork, "An Epitaxial Emitter Cap SiGe-Base Bipolar Technology for Liquid-Nitrogen Temperature Operation," *IEEE Electron Device Letters*, vol. 15, pp. 472-474, 1994.
- [83] P. Merken, T. Souverijns, J. Putzeys, Y. Creten, and V. Van Hoof, "Low noise, low power readout electronics circuit development in standard CMOS technology for 4 K applications," *Proc. SPIE*, vol. 6275, no. 6284, pp. 627 516-1-8, Jul. 2006.
- [84] H. Nagata, J. Kobayashi, H. Matsuo, and M. Fujiwara, "Progress on GaAs cryogenic readout circuits for SISCAM," *Proc. SPIE*, vol. 6275, no. 6284, pp. 627 527-1-10, Jul. 2006.
- [85] H. G. Liu, S. P. Walkins and C. R. Bolognesi, "15-nm base type-II InP/GaAsSb/InP DHBTs with  $f_T = 384$  GHz and a 6-V  $BV_{CEO}$ ," *IEEE Transactions on Electron Devices*, pp. 559 - 561, 2006.
- [86] H. G. Liu, O. Ostinelli, Y. Zeng and C. R. Bolognesi, "600 GHz InP/GaAsSb/InP DHBTs grown by MOCVD with Ga(AsSb) grade base and  $f_T \times BV_{CEO} > 2.5$  THz at room temperature," in *Tech. Dig. IEEE International Electron Devices Meeting*, pp. 667 - 670, 2007.
- [87] M. H. Khater, T. N. Adam, R. Krishnasamy, M. E. Dahlstrom, J.-S. Rieh, K. T. Schonenberg, B. A. Orner, F. Pagnette, K. Stien, and D. C. Ahlgren, "Present status and future directions of SiGe HBT technology," in *Proc. Lester Eastman Conference*, pp. 61-80, Aug. 2006.
- [88] N. Zerounian *et al.*, "Transit times of SiGe:C HBTs using nonselective base epitaxy," *Solid-State Electronics*, vol. 48, pp. 1993-1999, 2004.
- [89] P. Agarwal, G. A. M. Hurkx, J. J. T. M. Donkers, and J. W. Slotboom, "On the delay times in vertically scaled SiGe HBTs," in *Proc. IEEE Bipolar/BiCMOS Circuits Technology Meeting*, pp. 264-267, 2005.
- [90] M. Rodwell, E. Lind, Z. Griffith, S. R. Bank, A. M. Crook, U. Singiseti, M. Wistey, G. Burek, A. C. Gossard, "On the feasibility of few-THz bipolar transistors," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 18-21, 2007.

- [91] W. Lee *et al.*, "Monte Carlo simulation on non-equilibrium transport in ultra-thin base bipolar transistors," in *Tech. Dig. IEEE Int. Electron Devices Meeting*, pp. 473-476, 1989.
- [92] A.F.J. Levi, B. Jalali, R. N. Nottenburg, and A. Y. Cho, "Vertical scaling in heterojunction bipolar-transports with nonequilibrium base transport," *Applied Physics Letters*, vol. 60, no. 4, pp. 460-462, 1992.
- [93] M. A. Stettler and M. S. Lundstrom, "A microscopic study of transport in thin base silicon bipolar-transistors," *IEEE Transactions on Electron Devices*, vol. 41, no. 6, pp. 1027-1033, 1994.
- [94] D. M. Richey *et al.*, "Evidence for non-equilibrium base transport in Si and SiGe bipolar transistors at cryogenic temperatures," *Solid-State Electronics*, vol. 39, no. 6, pp. 785-789, June 1996.
- [95] B. K. Ridley and T. B. Watkins, "The possibility of negative resistance effects in semiconductors," *Proceedings of the Physical Society*, vol. 78, pp. 293-304, August 1961.
- [96] C. Hilsum, "Transferred electron amplifiers and oscillators," *Proc. IRE*, vol. 50, pp. 185-189, February 1962.
- [97] D. E. McCumber and A. G. Chynoweth, "Theory of negative-conductance amplification and of Gunn instabilities in 'two-valley' semiconductors," *IEEE Transactions on Electron Devices*, vol. ED-13, January 1966.
- [98] B. K. Ridley and T. B. Watkins, "The dependence of capture rate on electric field and the possibility of negative resistance in semiconductors," *Proceedings of the Physical Society*, vol. 78, pp. 710-715, November 1961.
- [99] B. K. Ridley and R. G. Pratt, "A bulk differential negative resistance due to electron tunneling through an impurity potential barrier," *Physics Letters*, vol. 4, pp. 300-302, May 1963.
- [100] J.J. ODwyer, "Theory of double charge ejection from a dielectric," *Journal of Applied Physics*, vol. 39, pp. 4356-4359, August 1968.
- [101] L. Esaki, "New phenomenon in narrow germanium P-N junctions," *Physical Review*, vol. 109, pp. 603-604, January 1958.
- [102] L. L. Chang, L. Esaki, R. Tsu, "Resonant tunneling in semiconductor double barriers," *Applied Physics Letters*, vol. 24, pp. 593-595, June 1974.
- [103] D. H. Chow, T. C. McGill, I. K. Sou, J. P. Faurie, and C. W. Nieh, "Observation of negative differential resistance from a single barrier heterostructure," *Applied Physics Letters*, vol. 52, pp. 54-56, January 1988.

- [104] H. Munekata, T. P. Smith, and L. L. Chang, "Growth and transport properties of (Ga,Al)Sb barriers on InAs," *Journal of Vacuum Science and Technology B (Microelectronics Processing and Phenomena)*, vol. 72, pp. 324-326, 1989.
- [105] J. R. Soderstrom, D. H. Chow, and T.C. McGill, "New negative resistance device based on resonant interband tunneling," *Applied Physics Letters*, vol. 55, pp. 1094-1096, November 1989.
- [106] P. See, D. J. Paul, B. Hollander, S. Mantl, I. V. Zozoulenko, and K. F. Berggren, "High performance Si/Si<sub>1-x</sub>Ge<sub>x</sub> resonant tunneling diodes," *IEEE Electron Device Letters*, vol. 22, pp. 182-184, 2001.
- [107] F. Capasso, S. Sen, F. Beltram, L. M. Lunardi, A. S. Vengurlekar, P. R. Smith, N. J. Shah, R. J. Malik, and A. Y. Cho, "Quantum functional devices: resonant-tunneling transistors, circuits with reduced complexity, and multiple-valued logic," *IEEE Transactions on Electron Devices*, vol. 36, pp. 2065-2082, 1989.
- [108] A. Cidronali, G. Collodi, M. Camprini, V. Nair, G. Manes, J. Lewis, and H. Goronkin, "Ultra low-power VCO based on InP-HEMT and heterojunction interband tunnel diode for wireless application," in *Tech. Dig. IEEE RFIC Symposium*, pp. 297-300, 2002.
- [109] T.P.E. Broekaert, B. Brar, J.P.A. van der Wagt, C. Seabaugh, T.S. Moise, and F.J. Morris, "A monolithic 4-bit 2-Gsps resonant tunneling analog-to-digital converter," *IEEE Journal of Solid State Circuits*, vol. 33, pp. 1342-1349, 1998.
- [110] S.J. Wei, H.C. Lin, R.C. Potter, and D. Shupe, "A self-latching A/D converter using resonant tunneling diodes," *IEEE Journal of Solid State Circuits*, vol. 28, pp. 697-700, 1993.
- [111] A. Miura, T. Yakihara, S. Uchida, S. Oka, S. Kobayashi, H. Kamada, "Monolithic sampling head IC," *IEEE Transactions on Microwave Theory and Technology*, vol. 38, pp. 1980-1985, 1990.
- [112] K. Arai, H. Matsuzaki, K. Maezawa, T. Otsuji, T. Yamamoto, "Static frequency divider featuring reduced circuit complexity by utilizing resonant tunneling diodes in combination with HEMTs," *IEEE Electron Device Letters*, vol. 18, pp. 544-546, 1997.
- [113] S.-Y. Chung, S. -Y. Park, J. W. Daulton, R. Yu, P.R. Berger and P.E. Thompson, "Integration of Si/SiGe HBT and Si-based RITD demonstrating controllable negative differential resistance for wireless applications," *Solid-State Electronics*, vol. 50, pp. 973-978, 2006.
- [114] R. Duane, A. Mathewson, and A. Concannon, "Bistable gated bipolar device," *IEEE Electron Device Letters*, vol. 24, pp. 661-663, 2003.

- [115] X. Cheng and R. Duane, "A comprehensive study of bistable gated bipolar device," *IEEE Transactions on Electron Devices*, vol. 53, no. 10, pp. 2589, 2006.
- [116] K. Ismail, B. S. Meyerson, and P. J. Wang, "Electron resonant tunneling in Si/SiGe double barrier diodes," *Applied Physics Letters*, vol. 59, pp. 973-975, 2000.
- [117] M. W. Dashiell, R. T. Troeger, S. L. Rommel, T. N. Adam, P. R. Berger, C. Guedj, J. Kolodzey, A. C. Seabaugh, R. Lake, "Current-voltage characteristics of high current density silicon Easki diodes grown by molecular beam epitaxy and the influence of thermal annealing," *IEEE Transactions on Electron Devices*, vol. 47, pp. 1707-1714, 2000.
- [118] E.O. Johnson, "Physical limitations on frequency and power parameters of transistors," *RCA Review*, vol. 26, pp. 163-177, June 1965.
- [119] J.-S. Rieh *et al.*, "A doping concentration-dependent upper limit of the breakdown voltage-cutoff Frequency product in Si bipolar transistors," *Solid State Electronics*, vol. 48, pp. 339-343, Feb. 2004.
- [120] T. Ishibashi *et al.*, "A possible near-ballistic collection in an AlGaAs/GaAs HBT with modified collector structure," *IEEE Transactions on Electron Devices*, vol. 35, No. 4, pp. 401-404, April 1988.
- [121] Y. Tateno *et al.*, "3.5V, 1W High Efficiency AlGaAs/GaAs HBTs with Collector Launcher Structure," in *Tech. Dig. IEEE International Electron Devices Meeting*, pp. 195-198, Dec. 1994.
- [122] T. Fujishira *et al.*, "Theory of Semiconductor Superjunction Devices," *Japanese Journal of Applied Physics*, vol. 36, pp. 6254-6262, Oct. 1997.
- [123] D. J. Coe, "High voltage semiconductor device," US Patent 4,754,310, issued 1988.
- [124] X. Chen, "Semiconductor power devices with alternating conductivity type high-voltage breakdown region," US Patent 5,216,275, issued 1993.
- [125] K. Ueno, "Super-junction semiconductor device," US Patent 6,512,268 B1, issued 2003.
- [126] D. M. Kinzer and S. Sridevan, "Lateral conduction superjunction semiconductor device," US Patent 6,787,872 B2, issued 2004.
- [127] M. Yamaguchi, W. Saito, I. Omura, and M. Izumisawa, "Semiconductor device with super junction region," US Patent 6,844,592 B2, issued 2005.
- [128] J.-G. Lee, J.-Y. Jung, and H.-C. Jang, "Superjunction semiconductor device," US Patent 7,301,203 B2, issued 2007.

- [129] J. D. Beasom, LBipolar method and structure having improved  $BV_{CEO} / R_{CS}$  trade-off made with depletable collector columns, U.S. Patent Application 2007/0273006 A1, filed 2007.
- [130] Mark Lundstrom, *Fundamentals of Carrier Transport*, Second Edition. Cambridge, UK: Cambridge University Press, 2000.
- [131] Carlo Jacoboni and Paolo Lugli, *The Monte Carlo Method for Semiconductor Device Simulation*. New York, NY: Springer-Verlag, 1989.
- [132] M. P. Anantram, M. S. Lundstrom, and D. E. Nikonov, "Modeling of nanoscale devices," *Proceedings of the IEEE*, vol. 96, no. 9, pp. 1511-1550, Sep. 2008.
- [133] G. Baccarani and M.R. Woerdeman, "An investigation of steady-state velocity overshoot in silicon," *Solid-State Electronics*, vol. 28, pp. 407-416, 1985.
- [134] H. Kroemer, "Two integral charge control relations pertaining to the electron transport through a bipolar transistor with a non-uniform energy gap in the base," *Solid-State Electronics*, vol. 28, pp. 1101-1103, 1985.
- [135] C. Jacoboni and L. Reggiani, "The Monte Carlo method for the solution of the charge transport in semiconductors with applications to covalent materials," *Reviews of Modern Physics*, vol. 55, no. 3, pp. 645-705, July 1983.
- [136] P. R. Gray, P. J. Hurst, S. H. Lewis, and R. G. Meyer, *Analysis and design of analog integrated circuits*, Fourth Edition, John Wiley & Sons, 2001.
- [137] R. C. Jaeger, *Microelectronic Circuit Design*, McGraw-Hill, 1997.
- [138] R. C. Jaeger and A. J. Brodersen, "Self consistent bipolar transistor models for computer simulation," *Solid-State Electronics*, vol. 21, pp. 1269-1272, 1978.
- [139] A. J. Joseph, J. D. Cressler, D. M. Richey, R. C. Jaeger, and D. L. Harame, "Neutral base recombination and its influence on the temperature dependence of Early voltage and current gain-Early voltage product in UHV/CVD SiGe heterojunction bipolar transistors," *IEEE Transactions on Electron Devices*, vol. 44, pp. 404-418, 1997.
- [140] J.D. Cressler, D.D. Tang, K.A. Jenkins, G.-P. Li, and E.S. Yang, "On the low-temperature static and dynamic properties of high-performance silicon bipolar transistors," *IEEE Transactions on Electron Devices*, vol. 36, no. 8, pp. 1489-1502, Aug. 1989.
- [141] J.M.C. Stork, D.L. Harame, B.S. Mayerson, and T.N.Nguyen, "Base profile design for high-performance operation of bipolar transistors at liquid-nitrogen temperature," *IEEE Transactions on Electron Devices*, vol. 36, no. 8, pp. 1503-1509, Aug. 1989.

- [142] N. Mott, "On the transition to metallic conduction in semiconductors," *Canadian Journal of Physics*, vol. 34, pp. 1356-1368, 1956.
- [143] E. Conwell, "Properties of silicon and germanium: II," *Proceedings of the IRE*, vol. 46, no. 6, pp. 1281-1300, June 1958.
- [144] A. Jonscher, "Semiconductors at cryogenic temperatures," *Proc. IEEE*, pp. 1092-1104, Oct. 1964.
- [145] P.P. Altermatt, A. Schenk, B. Schmithusen, and G. Heiser, "A simulation model for the density of states and for incomplete ionization in crystalline silicon. II. Investigation of Si:As and Si:B and usage in device simulation," *Journal of Applied Physics*, vol. 100, no. 11, pp. 113715, 2006.
- [146] W.P. Dumke, "Effect of minority carrier trapping on the low-temperature characteristics of Si transistors," *IEEE Transactions on Electron Devices*, vol. 17, no. 4, pp. 388-389, Apr. 1970.
- [147] K. A. Moen and J. D. Cressler, "Measurement and modeling of carrier transport parameters in SiGe BiCMOS technology operating in extreme environments," *IEEE Transactions on Electron Devices*, in press.
- [148] Y. Cui, G. Niu, Y. Shi, C. Zhu, L. Najafizadeh, J.D. Cressler, A. Joseph, "SiGe Profile Optimization for Improved Cryogenic Operation at High Injection", in *Proc. Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 170-173, Oct. 2006.
- [149] H. Satake and T. Hamasaki, "Low-temperature (77 K) BJT model with temperature dependences on the injected condition and base resistance," *IEEE transactions on Electron Devices*, vol. 37, no. 7, pp. 1688-1697, July 1990.
- [150] E. A. Gutierrez-D., M. -J. Deen, and C. L. Claeys, *Low Temperature Electronics: Physics, Devices, Circuits, and Applications*, Academic Press, London, 2001.
- [151] W.-M. L. Kuo, Q. Liang, J. D. Cressler, and M. A. Mitchell, "An X-band SiGe LNA with 1.36 dB mean noise figure for monolithic phased array transmit/receive radar modules." in *Tech. Dig. Radio Frequency Integrated Circuits Symposium*, pp. 498-501, 2006.
- [152] M. W. Pospieszalski, "Extremely low noise amplification with cryogenic FETs and HEMTs: 1970-2004," *IEEE Microwave Magazine*, Vol. 6, pp. 62-75, 2005.
- [153] J. Kraft, D. Kraft, B. Loffler, H. Jauk, and E. Wachmann, "Usage of HBTs beyond  $BV_{CEO}$ ", in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 33-36, 2005.
- [154] C. M. Grens, J. D. Cressler, J. M. Andrews, Q. Liang, and A. J. Joseph, "The effects of scaling and bias configuration on operating-voltage constraints in SiGe HBTs for mixed-signal circuits," *IEEE Trans. Electron Devices*, vol. 54, no. 7, 2007.

- [155] H. Rucker, B. Heinemann, W. Winkler, R. Barth, J. Borngraber, J. Drews, G. G. Fischer, A. Fox, T. Grabolla, U. Haak, D. Knoll, F. Korndorfer, A. Mai, S. Marschmeyer, P. Schley, D. Schmidt, J. Schmidt, K. Schulz, B. Tillack, D. Wolansky, and Y. Yamamoto, "A 0.13m SiGe BiCMOS technology featuring  $f_T/f_{max}$  of 240/330 GHz and gate delays below 3 ps," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, paper 11.1, Oct. 2009.
- [156] C. J. Glassbrenner and G. A. Slack, "Thermal conductivity of silicon and germanium from 3 K to the melting point," *Phys. Rev.*, vol. 134, no. 4A, pp. A1058-A1069, May 1964.
- [157] J.C. Thompson and B.A. Younglove, "Thermal conductivity of silicon at low temperatures," *J. Phys. Chem. Solids*, vol. 20, no. 1/2, pp. 146-149, 1961.
- [158] L. Weber and E. Gmelin, "Transport properties of silicon," *Applied Physics A: Materials Science and Processing*, vol. 53, no. 2, pp. 136-140, 1991.
- [159] C. Raya, F. Pourchon, M. Laurens, and T. Zimmer, "New test structures for extraction of base sheet resistance in BiCMOS technology," in *Proc. IEEE International Conference on Microelectronic Test Structures*, pp. 35-40, March 2006.
- [160] D.B.M. Klaassen, "A Unified Mobility Model for Device Simulation-I. Model Equations and Concentration Dependence," *Solid-State Electronics*, vol. 35, pp. 953-959, 1992.
- [161] D.B.M. Klaassen, "A Unified Mobility Model for Device Simulation-II. Temperature Dependence of Carrier Mobility and Lifetime," *Solid-State Electronics*, vol. 35, pp. 961-967, 1992.
- [162] C.R. Battjes, "Monolithic wideband amplifier," US Patent 4,236,119, Nov. 1980.
- [163] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, 2nd ed., Cambridge, UK: Cambridge Univ. Press, 2004.
- [164] C. Hermans and M.S.J. Steyaert, "A high-speed 850-nm optical receiver front-end in 0.18- $\mu$ m CMOS," *IEEE Journal of Solid State Circuits*, pp.1606, 2006.
- [165] M. Joodaki, "Small-signal characterization of SiGe-HBT  $f_T$ -doubler up to 120 GHz," *IEEE Transactions on Electron Devices*, pp. 2108-2111, 2005.
- [166] M Chu, R. Heikaus, K. Zhou, J.-R. Gun, C. You, J.F. McDonald, P.P. Kraft, "Ultra high speed interleaved aid conversion using an  $f_T$  doubler core in SiGe HBT technology," in *Proc. IEEE Instrumentation and Measurement Technology Conference*, vol. 1, pp. 839-844, May 2003.
- [167] J.-L. Olvera-Cervantes, J.D. Cressler, J.-L. Medina-Monroy, T. Thiruvikraman, B. Banerjee, J. Laskar, "A new analytical method for robust extraction of the small-signal equivalent circuit for SiGe HBTs operating at cryogenic temperatures," *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, no. 3, pp. 568-574, March 2008.



- [168] P. M. Solomon and D. D. Tang, "Bipolar circuit scaling" in *Dig. Int. Solid-State Circuits Conf.*, pp. 86-87, 1979.
- [169] G. P. Rosseel and R. W. Dutton, "Scaling rules for bipolar transistors in BiCMOS circuits," in *Tech. Dig. Int. Electron Devices Meeting*, pp. 795-798, 1989.
- [170] A. Bellaouar, S. H. K. Embabi, and M. I. Elmasry, "Scaling of digital BiCMOS circuits," *IEEE J. Solid-State Circuits*, vol. 25, no. 4, pp. 932-941, Aug. 1990.
- [171] P. A. Raje, K. C. Saraswat, and K. M. Cham, "Performance-driven scaling of BiCMOS technology," *IEEE Trans. Electron Devices*, vol. 39, no. 3, pp. 685-694, Mar. 1992.
- [172] K. K. Ng, M. R. Frei, and C. A. King, "Reevaluation of the  $f_tBV_{ceo}$  limit on Si bipolar transistors," *IEEE Trans. Electron Devices*, vol. 45, no. 8, pp. 1854 - 1855, Aug. 1998.
- [173] B. Heinemann, R. Barth, D. Bolze, J. Drews, P. Formanek, T. Grabolla, U. Haak, W. Hoppner, D. Knoll, K. Kopke, B. Kuck, R. Kurps, S. Marschmeyer, H. Richter, H. Rucker, P. Schley, D. Schmidt, W. Winkler, D. Wolansky, H. -E. Wulf, and Y. Yamamoto, "A low-parasitic collector construction for high-speed SiGe:C HBTs," in *Tech. Dig. Int. Electron Devices Meeting*, pp. 251-254, 2004.
- [174] J. J. T. M. Donkers, M. C. J. C. M. Kramer *et al.*, "A novel fully self-aligned SiGe:C HBT architecture featuring a single-step epitaxial collector-base process," in *Tech. Dig. Int. Electron Devices Meeting*, pp. 655-658, 2007.
- [175] J. J. T. M. Donkers, T. Vanhoucke, P. Agarwal, R. J. E. Hueting, P. Meunier-Beillard, M. N. Vijayaraghavan, P. H. C. Magnee, M. A. Verheijen, R. de Kort, and J. W. Slotboom, "Metal emitter SiGe:C HBTs", in *Tech. Dig. Int. Electron Devices Meeting*, pp. 243 - 246 Dec. 2004.
- [176] M. Xu, S. Decoutere, A. Sibaja-Hernandez, K. Van Wichelen, L. Witters, R. Loo, *et al.*, "Ultra low power SiGe:C HBT for 0.18  $\mu\text{m}$  RF-BiCMOS," in *Tech. Dig. Int. Electron Devices Meeting*, pp. 125-128, 2003.
- [177] DESSIS 10 User Manual. Zurich, 2004.
- [178] H. Nakajima, M. Tomizawa, and T. Ishibashi, "Monte Carlo analysis of the space-charge effect in AlGaAs/GaAs ballistic collection transistors (BCTs) under high current injection," *IEEE Transactions on Electron Devices*, vol. 39, no. 7, pp. 1558-1563, July 1992.
- [179] T. Kurosawa, "Monte Carlo calculation of hot-electron problems," in *Proceedings of the International Conference on the Physics of Semiconductors, Kyoto, J. Phys. Soc. Jap. Suppl.* vol. 21, pp. 424-426, 1966.
- [180] Sentaurus Device Monte Carlo User Guide, Synopsys, September 2008.

- [181] P. Palestri, A. Pacelli, M. Mastrapasqua, and J. D. Bude, "Monte Carlo simulation of impact ionization in SiGe HBTs," *IEEE Electron Device Letters*, vol. 22, pp. 533-535, Nov. 2001.
- [182] S. Keith, C. Jungemann, S. Decker, B. Neinhuis, M. Bartels, B. Meinerzhagen, "Full-band Monte Carlo device simulation of a Si/SiGe-HBT with a realistic Ge profile," in *Proc. International Conference on Simulation of Semiconductor Processes and Devices*, pp. 219-222, 1999.
- [183] A. Leone, A. Gnudi, and G. Baccarani, "Hydrodynamic simulation of semiconductor devices operating at low temperature," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 13, no. 11, Nov. 1994.
- [184] P. Palestri, L. Selmi, G.A.M. Hurkx, J.W. Slotboom, D. Terpstra, M. Peter, R. Woltjer, E. Sangiorgi, "Non Local Electron and Hole Impact Ionization in Advanced Si BJTs," *Philips Research Laboratory, Eindhoven, The Netherlands*, Feb. 2000.
- [185] R.F. Pierret, *Semiconductor Device Fundamentals*, Reading, MA: Addison-Wesley, 1996.
- [186] S.E. Laux and M.V. Fischetti, "Monte-Carlo simulation of submicrometer Si n-MOSFET's at 77 and 300 K," *IEEE Electron Device Letters*, vol. 38, no. 6, pp. 1343-1352, 1991.
- [187] S.-M. Hong and C. Jungermann, "Electron Transport in extremely scaled SiGe HBTs," in *Proc. IEEE Bipolar/BiCMOS Circuit and Technology Meeting*, pp. 67, Oct. 2009.
- [188] Carrie Ann Niemiec and Mark S. Lundstrom, "Benchmarking and verification of bipolar junction transistor TCAD simulation Programs," ECE Technical Reports, Purdue Libraries, 1996. Available at <http://docs.lib.purdue.edu/ecetr/94>.

## VITA

Jiahui Yuan was born in Shanghai in 1983. He received his B.Eng. degree in 2005 from the Department of Electronic Engineering, Tsinghua University, Beijing, China, and the M.S. degree in Electrical Engineering from Georgia Institute of Technology (Georgia Tech), Atlanta, GA, USA, in 2007. His master's thesis was *SiGe HBTs Operating at Deep Cryogenic Temperatures*. He obtained the highest GPA in his class for his undergraduate studies and maintains a perfect 4.0 GPA at graduate school.

As an undergraduate research assistant in the CAD group of the Institute of Microelectronics, Tsinghua University, he worked on the simulation and measurement of the thermal conductivity of polysilicon ultra-thin films during his junior year (2004) under the direction of a graduate student. In his senior year, he focused on the design and optimizations of 90 nm strained-Si PMOS transistors by using process and device simulators on his own.

Yuan joined Professor John D. Cressler's silicon-germanium (SiGe) team in the School of Electrical and Computer Engineering, Georgia Tech, Atlanta in 2005. He received an *IEEE Electron Device Society Ph.D. Student Fellowship Award* in 2008, and he is a reviewer of *IEEE Electron Device Letters*, *Cryogenics*, and *IEEE Transactions on Electron Devices*.

Yuan has been serving in Habitat for Humanity in Atlanta since 2007.